REMARKS

This Reply is responsive to the non-final Office Action dated March 8, 2007. Entry of the amendments and remarks submitted herein and reconsideration of the claimed subject matter pursuant to 37 CFR §1.111 is respectfully requested.

I. Status of the Claims

Claims 68-173 were pending in this application at the time of the Office Action dated March 8, 2007. Claims 68-106, 169 and 170 were withdrawn from consideration pursuant to a restriction requirement, but have not been cancelled. The Office Action Summary does not refer to the withdrawn claims, therefore, clarification is respectfully requested. Claims 107-168 and 171-173 were under examination at the time of the Office Action dated March 8, 2007. New claim 174 is added by way of amendment above. Therefore, claims 68-174 are currently pending, with claims 107-168 and 171-174 currently under examination.

II. Claim Amendments

Claims 107 and 137 were amended to clarify that the target gene is a mammalian gene or a mammalian pathogen gene, or found in a mammalian cell. Support for this amendment may be found at page 7, lines 4-9.

Claim 139 was amended above to indicate that the claimed expression vector may encode two or more RNA molecules that may be at least about 100 nucleotides in length and comprise at least 11 to 30 nucleotides involved in the double-stranded sequence. Support for this amendment may be found at page 8, lines 24-25, and page 9, lines 5-18, of the specification. New claim 174 was added which indicates that the claimed expression vector may encode two or

more RNA molecules that may be less than about 750 nucleotides in length and comprise at least 11 to 30 nucleotides involved in the double-stranded sequence. Support for this amendment may be found at page 9, lines 3-4, and page 9, lines 5-18, of the specification. No prohibited new matter has been added by way of this amendment.

III. Priority Determination

Applicants acknowledge with appreciation the indication at page 2 of the Office Action that claims 107-138, 140-168 and 171-173 have been accorded a priority date of 4/19/2000. Claim 139, however, has only been accorded a priority date of 10/2/2002, the filing date of the instant application, for the reasons set forth in the new matter rejection. Applicants have addressed the arguments raised in the new matter rejection below, and believe that claim 139 as amended is fully supported by the specification as filed. Accordingly, Applicants respectfully request that claim 139, as well as new claim 174, be accorded at least the priority date 4/19/2000, the filing date of the parent PCT application.

IV. Rejection under 35 USC §112

Claim 139 has been rejected under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement. According to the Office Action, this is a new matter rejection. Essentially, while the Examiner acknowledges that the specification discloses that the entire length of the RNA molecule of the invention can be double stranded, the Examiner also asserts that the RNA polynucleotide must be at a minimum 100 nucleotides in length (Office Action, page 4). Therefore, the Examiner asserts that the specification does not contemplate an expression vector encoding two or

more different double stranded sequences wherein each different double stranded sequence comprises only 11 to 30 nucleotides involved in the double stranded sequence. Applicants respectfully traverse the rejection.

Applicants respectfully note that the length range to which the Examiner refers is only a *preferred* range. Indeed, the specification states that "The 'at least partially double stranded RNA molecule' *includes* an RNA polynucleotide sequence of between about 100 to 10,000 polynucleotides in length" (page 8, lines 23-24, with emphasis). Nevertheless, without agreeing with the rejection, Applicants have amended claim 139 above to indicate that the vector may encode two or more RNA molecules that are at least about 100 nucleotides in length and comprise at least 11 to 30 nucleotides involved in the double-stranded sequence. Given that claim 139 as amended has explicit support in the application as filed, Applicants respectfully request reconsideration and withdrawal of this rejection under §112, first paragraph.

V. Prior Art Rejections

Claim 139 was rejected under 35 U.S.C. §102(a) as being anticipated by Leirdal *et al.* (2002). According to the Office Action, Leirdal *et al.* teaches a multitarget partially double stranded RNA molecule comprising two different double stranded RNA sequences that are complementary to a GFPsi1 sequence and a PKCαsi3 sequence, respectively. Applicants respectfully traverse the rejection.

As Applicants have argued above in response to the rejection under §112, first paragraph, claim 139 as amended has explicit support in the instant specification. The instant specification is identical to the specification of the parent PCT application filed

4/19/2000. Accordingly, claim 139 should be granted the priority date of the PCT application (4/19/2000), and Leirdal *et al.* is not prior art as it was published in 2002. Reconsideration and withdrawal of the rejection of claim 139 under §102(a) based on Leirdal *et al.* are respectfully requested.

Claims 107-168 and 171-173 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Werther *et al.* (US 5,929,040), Fire *et al.* (US 6,506,559), Heifietz *et al.* (WO 99/61631), Calabretta *et al.* (US 5,734,039) and Thompson *et al.* (US 6,146,886). According to the Office Action, Werther *et al.* teaches a multivalent antisense molecule but does not disclose the use of double stranded RNA sequences or the expression of double stranded RNA sequences from a vector. Calabretta *et al.* allegedly teaches a vector for the expression of a first and second antisense molecule under the control of first and second promoters, respectively, but like Werther, also does not teach the expression of double stranded RNA sequences. However, the Examiner believes it would have been obvious to the skilled artisan at the time the invention was made to substitute double stranded RNA as allegedly disclosed in Fire *et al.* and Heifetz *et al.* as an alternative for the antisense sequences in the constructs of Werther and Calabretta and express these molecules in a vector as allegedly also disclosed in Fire *et al.* and Heifetz *et al.* Thompson is relied upon for teaching expression of therapeutic RNAs including ribozymes and antisense RNAs using a RNA polIII promoter. Applicants respectfully traverse the rejection.

There are two general embodiments covered in the instant claims. First, the instant claims as amended encompass multitarget partially double stranded RNA

molecules comprising two or more different double stranded RNA sequences that are substantially homologous and complementary to two or more sequences of at least one target mammalian gene or mammalian pathogen gene (claim 107), and vectors expressing such multitarget constructs (i.e., claim 141). The instant claims also encompass expression vectors encoding multiple double stranded RNA sequences that are expressed from different promoters (i.e., claim 150). Neither of these general embodiments is rendered obvious by the cited references.

With reference to the first general embodiment, it would not have been obvious to the skilled artisan to combine Werther and Fire to produce a multitarget double stranded RNA molecule as claimed. At the time the invention was made, it was widely presumed that long double stranded RNA molecules cause an interferon stress response or "PKR response" when introduced into mammalian cells. For instance, as discussed in Manche et al. (1992, Mol. Cell Biol. 12(11): 5238-48) (copy attached), it was reported that double stranded RNA molecules longer than 30 base pairs activate interferon-induced protein kinase. Bevilacqua and colleagues (1998, Biochem. 37(18): 6303-16) (copy attached) also reported that interferon-induced PKR is activated by long stretches of dsRNA (see abstract), and in particular, that binding sites for PKR "may be assembled from several noncontiguous helices" (p. 6314, col. 1).

Accordingly, the skilled artisan would not have been motivated to produce long multitarget *double stranded* RNA molecules that target mammalian genes since, *unlike the situation with the antisense construct proposed in Werther*, the skilled artisan would have wanted to avoid a PKR response. Indeed, the tendency at the time of the present invention was to use *shorter* dsRNA molecules for this very reason, not longer

multitarget molecules as encompassed by the present claims. See, *e.g.*, Elbashir et al. (2001, Genes & Dev. 15: 188-200) (copy attached).

In view of the references discussed above, it is clear that the scientific community presumed at the time the invention was made that longer double stranded RNA molecules and double stranded RNAs with partial secondary structures would invoke a PKR response in mammalian cells. Therefore, there would have been no expectation of success in producing a multitarget partially double stranded RNA molecule targeting two or more mammalian gene sequences. Indeed, evidence showing that there was no reasonable expectation of success supports a conclusion of nonobviousness. *In re Rinehart*, 531 F.2d 1048, 189 USPQ 143 (CCPA 1976).

With regard to the second general embodiment, i.e., expression vectors encoding multiple double stranded RNA sequences that are expressed from different promoters, the skilled artisan would not have been motivated to express the double stranded RNA sequences of the invention from separate promoters on the same vector due to the well known phenomenon of promoter or transcriptional interference. At the time the present invention was made, it was well known in the art that competitive interference between promoters on a single vector commonly results in preferential expression of one gene over the other(s). See, for example, Ghattas et al. (1991, Mol. Cell. Biol. 11(12): 5848-59); Morgan et al. (1992, Nuc. Acids Res. 20(6): 1293-99); and Chen et al. (1993, J. Virol. 67(4): 2142-48) (copies attached). In fact, according to Chen et al., they observed that an exogenous gene expressed from a double-gene vector "is never expressed as efficiently as from a single gene vector" (p. 2142, col. 1).

Given what was known of promoter interference at the time the invention was made, the skilled artisan would not have been motivated to express more than one double stranded RNA from separate promoters on a single vector. In contrast to what would have been expected in view of the state of the art, Applicants have surprisingly found that different double stranded RNAs may be co-expressed efficiently from the same vector. In this regard, Applicants note that although Calabretta suggests that two antisense sequences may be expressed on a single vector construct from separate promoters, Applicants also observe that there is no indication in the Calabretta disclosure that this was actually performed.

In summary, the skilled artisan would not have been motivated to produce long multitarget double stranded RNA molecules that target mammalian genes since, unlike the situation with the antisense construct proposed in Werther, the skilled artisan would have wanted to avoid a PKR response. Further, the skilled artisan would not have been motivated to express several different double stranded RNAs from a single vector, since it was commonly assumed in the prior art that promoter suppression or promoter interference complicated expression of dual constructs from a single vector. Thompson was only relied upon for teaching expression of therapeutic RNAs including ribozymes and antisense RNAs using a RNA polIII promoter. Thompson does not make up for the deficiencies of Werther, Calabretta, Fire and Heifetz with regard to teaching multitarget dsRNA molecules and vector constructs expressing multiple dsRNAs. In view of all the above remarks, reconsideration and withdrawal of the rejection under §103(a) are respectfully requested.

In summary, Applicants believe that this Reply adequately addresses all the rejections of the claims, and that the application should now be allowed. Applicants respectfully note that this is the third non-final Office Action mailed for the subject application. Therefore, favorable indication of allowable subject matter is respectfully requested.

Except for issue fees payable under 37 CFR §1.18, the commissioner is hereby authorized by this paper to charge any additional fees during the pendency of this application including fees due under 37 CFR §1.16 and 1.17 which may be required, including any required extension of time fees, or credit any overpayment to Deposit Account 50-1283. This paragraph is intended to be a **CONSTRUCTIVE PETITION FOR EXTENSION OF TIME** in accordance with 37 CFR §1.136(a)(3).

If the Examiner has any further questions relating to this Reply or to the application in general, she is respectfully requested to contact the undersigned by telephone so that allowance of the present application may be expedited.

Dated: June 8, 2007

Respectfully submitted,
COOLEY GODWARD KRONISH LLP

By: Mark Hayman #51,793
for Bonnie Weiss McLeod
Reg. No. 43,255

CUSTOMER NO.: 58249
COOLEY GODWARD KRONISH LLP
ATTN: Patent Group
1200 19th Street, NW, 5th Floor
Washington, DC 20036-2412
Tel: (202) 842-7833
Fax: (202) 842-7899

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1992, p. 5238–5248
0270-7306/92/115238-11$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 12, No. 11

Interactions between Double-Stranded RNA Regulators and the Protein Kinase DAI

LISA MANCHE, SIMON R. GREEN, CHRISTIAN SCHMEDT, AND MICHAEL B. MATHEWS*
Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, New York 11724

Received 6 May 1992/Returned for modification 2 July 1992/Accepted 27 August 1992

The interferon-induced protein kinase DAI, the double-stranded RNA (dsRNA)-activated inhibitor of translation, plays a key role in regulating protein synthesis in higher cells. Once activated, in a process that involves autophosphorylation, it phosphorylates the initiation factor eIF-2, leading to inhibition of polypeptide chain initiation. The activity of DAI is controlled by RNA regulators, including dsRNA activators and highly structured single-stranded RNAs which block activation by dsRNA. To elucidate the mechanism of activation, we studied the interaction of DAI with RNA duplexes of discrete sizes. Molecules shorter than 30 bp fail to bind stably and do not activate the enzyme, but at high concentrations they prevent activation by long dsRNA. Molecules longer than 30 bp bind and activate the enzyme, with an efficiency that increases with increasing chain length, reaching a maximum at about 85 bp. These dsRNAs fail to activate at high concentrations and also prevent activation by long dsRNA. Analysis of complexes between dsRNA and DAI suggests that at maximal packing the enzyme interacts with as little as a single helical turn of dsRNA (11 bp) but under conditions that allow activation the binding site protects about 80 bp of duplex. When the RNA-binding site is fully occupied with an RNA activator, the complex appears to undergo a conformational change.

Protein synthesis is modulated at several levels, most commonly at the stage of polypeptide chain initiation, and the phosphorylation of initiation factors plays a key role in controlling this process (reviewed in references 19 and 20). In mammalian cells, a regulatory mechanism involving an RNA-activated protein kinase and the eukaryotic initiation factor 2 (eIF-2) has been intensively studied. This initiation factor forms a ternary complex with GTP and Met-tRNA$_F$ and delivers the initiator tRNA to the ribosomal site of protein synthesis initiation. Discharged eIF-2 is subsequently released as a complex with GDP which must be replaced with GTP to permit the formation of another ternary complex in preparation for a further round of initiation. The factor is composed of three dissimilar subunits, α, β, and γ. Phosphorylation of a single residue, serine-51 of the α subunit, inhibits translation by trapping a second initiation factor, the guanosine nucleotide exchange factor (or eIF-2B), which is required to catalyze the substitution of GTP for GDP in the discharged eIF-2 complex. Phosphorylation of sufficient eIF-2 can sequester all of the guanosine nucleotide exchange factor, thereby preventing eIF-2 recycling and halting the initiation pathway.

In mammals, two protein kinases are capable of phosphorylating the α subunit of eIF-2 in this way (reviewed in references 20, 37, and 46). One of them, the heme-controlled repressor, is found chiefly in reticulocytes. It is activated by the absence of hemin, as well as by other stimuli, and serves to prevent the accumulation of globin in the absence of iron or heme. A second kinase, the double-stranded RNA-activated inhibitor (DAI; also referred to as P1 kinase, p68 kinase, P1/eIF-2α kinase, and PK$_{ds}$, etc.) is present in a wide range of tissues. DAI is an important element in the host antiviral response, and its synthesis is induced at the transcriptional level by interferon (reviewed in references 21, 54, 56, and 59). The enzyme is ribosome associated (11, 34) and normally exists in an inactive or latent state. Under some circumstances, DAI activation leads to the virtually complete abrogation of protein synthesis, while in other circumstances it may contribute to the selective translation of particular classes of mRNA (8, 24, 26, 36, 47, 60). It has also been implicated in cellular differentiation (23, 52), in the inhibition of cell proliferation (6, 51), in the heat shock response (10), and possibly in transcriptional induction (61, 64). Moreover, in yeast cells, the related protein kinase GCN2 mediates the growth response to amino acid starvation (9). As its name implies, DAI is activated by double-stranded RNA (dsRNA). Other polyanions such as heparin can also activate it, while small, highly structured RNA molecules such as adenovirus VA RNA suppress its activation (38). Thus, DAI is a pivotal cellular regulatory enzyme whose level and activity are modulated by factors of both viral and cellular origin.

The interactions between DAI and its RNA effectors are complicated and incompletely understood. The kinase is activated by dsRNA but not by DNA or DNA-RNA hybrids (22, 32, 35, 58). Single-stranded RNA, either synthetic or natural, is also inactive unless it can form extended hairpin-like structures (5, 22). There is no discernible sequence dependence for activation by dsRNA, and limited mismatching (44) and some modified bases (2, 45, 62) are tolerated, but the activity of dsRNA is reduced by ethidium bromide (1), suggesting that the topological form of the RNA duplex is important. Activation is accompanied by autophosphorylation of the kinase at multiple sites on serine and threonine residues (3, 11, 14, 30), and results in a change of substrate specificity such that the activated enzyme can phosphorylate the α subunit of eIF-2 and some other proteins (53, 58) but can no longer phosphorylate other molecules of DAI (29). Once activated, however, the phosphorylated enzyme is unaffected by the addition or removal of dsRNA (11, 58, 63).

Activation of DAI by dsRNA displays a paradoxical concentration dependence: the enzyme is activated by low concentrations of dsRNA (in the range of 10 to 100 ng/ml), but higher concentrations are decreasingly effective activators, giving rise to a bell-shaped activation curve (11, 22, 27,

* Corresponding author.

Downloaded from mcb.asm.org by on March 28, 2007

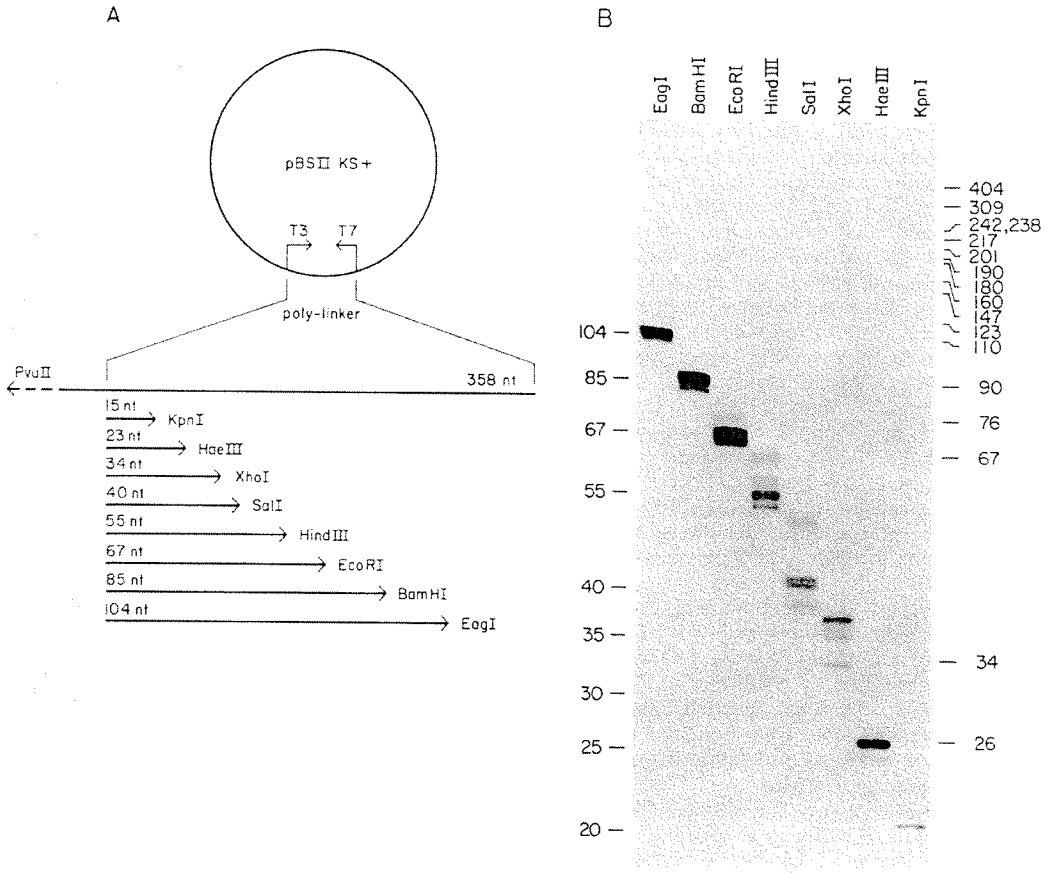


FIG. 1. Synthesis and characterization of RNA duplexes. (A) Schematic of short dsRNAs produced by transcription of pBSII KS+ polylinker sequences. The several transcripts of 15 to 104 nt synthesized by T3 RNA polymerase (rightward arrows) were annealed to the complementary 358-nt transcript (leftward arrows) synthesized by T7 RNA polymerase. After RNase digestion, the duplexes were purified by electrophoresis in nondenaturing gels. (B) Analysis of the purified dsRNAs in denaturing conditions. Samples of the radiolabeled dsRNAs were heated in formamide and resolved by electrophoresis in a 10% polyacrylamide–7 M urea gel. The fixed and dried gel was subjected to autoradiography. Size markers were single-stranded RNAs synthesized as described in part A (left) and pBR322-*Hpa*II DNA fragments (right).

33). High concentrations of dsRNA prevent the activation process but do not interfere with the activity of DAI once it has been activated by dsRNA at a lower concentration. The kinase also displays a stringent requirement for dsRNA chain length. Activation is reported to require a minimum of about 50 bp of duplex (18, 22, 32, 44), and there are indications that shorter duplexes may block activation at high concentration (44), as long dsRNA does. The ability of the enzyme to discriminate between dsRNA molecules on the basis of their chain length has implications for its regulation and the mechanism of DAI activation (37). Here we investigate the interactions of the enzyme with dsRNA molecules of specified sizes, studying binding and protection of dsRNA as well as activation and inhibition of the kinase. Our results define the dsRNA size dependence of the interaction and confirm that short duplexes which fail to bind stably and to activate the kinase can still interfere with activation mediated by longer duplexes. The data suggest that the minimal recognition element is a single helical turn but that there is an extended site for dsRNA binding which needs to be completely occupied for full enzyme activation.

MATERIALS AND METHODS

Synthesis of short dsRNAs. The plasmid pBSII KS+ (Stratagene, Inc., La Jolla, Calif.) was banded twice in CsCl, passed over a Bio-Gel A 15-m column, and then digested with one of eight enzymes (*Kpn*I, *Hae*III, *Xho*I, *Sal*I, *Hin*dIII, *Eco*RI, *Bam*HI, or *Eag*I), which cut in the polylinker, or with *Pvu*II, which cuts outside the region containing the polylinker and the T3 and T7 promoters (Fig. 1A). The DNA was incubated with RNase A to remove the last traces of RNA, treated with proteinase K, and extracted with phenol and chloroform. After ethanol precipitation, the DNA was added to transcription reactions containing T7 RNA polymerase (17) for the PvuII-digested template or T3 RNA polymerase (Stratagene, Inc.) for the other templates. Reaction conditions were as described previously (43), except that the concentration of GTP or CTP was reduced to 12 μM for labeling. The corresponding [α-^{32}P]ribonucleotide (from ICN Biomedicals Inc., Costa Mesa, Calif.) was present at a concentration of 500 μCi/ml. Single-stranded RNA was recovered after DNaseI digestion and phenol and chloroform extraction by ethanol precipitation. Each of the

Downloaded from mcb.asm.org by on March 28, 2007

T3 products (15 to 104 nucleotides [nt]) was mixed with an approximately equivalent amount of the complementary T7 product (354 nt), heated to 100°C, and annealed as previously described (29). Following digestion with both RNase T_1 and RNase A, dsRNA was isolated by treatment with proteinase K and deproteinization and then fractionated by electrophoresis in a nondenaturing 10% polyacrylamide–0.5× Tris-borate-EDTA (TBE) gel. The bands were detected autoradiographically, and each dsRNA was eluted into 10 mM Tris–1 mM EDTA –10 mM NaCl–0.5% sodium dodecyl sulfate (SDS), deproteinized, and ethanol precipitated. The dsRNA was dissolved in the same buffer without SDS, and its concentration was calculated from the specific activity.

Other RNAs. Longer dsRNAs (354 bp) were synthesized by transcription of the pGEM.GC plasmid (42). Reovirus dsRNA was provided by A. J. Shatkin (Rutgers University, New Brunswick, N.J.), and *Penicillium chrysogenum* and bacteriophage f2 sus3 dsRNAs were provided by H. D. Robertson (Cornell University Medical School, Ithaca, N.Y.). Labeled single-stranded RNA was purified from the T7 and T3 polymerase transcription reactions described above by electrophoresis through a 10% polyacrylamide–7 M urea–0.5× TBE gel.

Kinase assays. Reactions (10 μl) containing 2.5 μCi of [γ-^{32}P]ATP (ICN Biomedicals, Inc.) and 0.5 μl of DAI (about 5 ng) purified to the Mono S stage (29) were conducted essentially as described by Mellits et al. (42). The enzyme was added last to the other reaction components assembled on ice. Phosphorylation was visualized by SDS-polyacrylamide gel electrophoresis and autoradiography for 4 to 16 h by using an intensifier screen.

Nitrocellulose filter-binding assay. The nitrocellulose filter-binding assay was conducted by using a modification of the published procedure of Kostura and Mathews (29). Briefly, labeled dsRNA was incubated for 20 min on ice with the Mono S fraction of DAI under kinase reaction conditions, with bovine serum albumin (BSA) and calf liver tRNA both added to a concentration of 0.1 mg/ml but with labeled ATP omitted. After dilution with 10 volumes of wash buffer (50 mM KCl, 1.5 mM $MgCl_2$, 20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES) K^+ [pH 7.4], 0.1 mM EDTA), the reaction mixtures were immediately filtered in a slot-blot apparatus through a 0.45-μm-pore-size nitrocellulose membrane (Schleicher & Schuell, Keene, N.H.) that had been soaked for 1 h at room temperature in wash buffer containing 0.1 mg each of BSA and salmon sperm DNA per ml. Each well was washed with 200 μl of ice-cold wash buffer, and the filter was dried and exposed to autoradiography. Quantitation was done by scintillation counting of individual bands or direct scanning of the membrane with the AMBIS Imaging System.

Binding of dsRNA to Sepharose-bound DAI. A mixture of dsRNAs was partially degraded by incubation with RNase T_1 and RNaseIII (provided by H. D. Robertson) and then incubated with DAI immobilized on monoclonal antibody-Sepharose beads (13, 31) (from A. Hovanessian, Institut Pasteur, Paris, France) as described previously (40). The beads were sedimented and washed five times by resuspension and sedimentation. Of the input radioactivity, approximately 15% was recovered with the beads, 80% was recovered in the initial supernatant plus first wash fraction and 5% was recovered in subsequent washes. RNA was extracted from the beads and from the initial supernatant plus the first wash fraction and was analyzed by electrophoresis in a nondenaturing 10% polyacrylamide–0.5× TBE gel.

Protection of dsRNA by DAI. Radiolabeled dsRNA (354 bp) was bound to immobilized DAI as described above. After the third wash, the beads were washed twice with RNaseIII buffer (100 mM NH_4Cl, 10 mM magnesium acetate, 20 mM Tris-HCl [pH 7.6]) and then incubated with 50 U of RNaseIII per ml for 30 min at 37°C. An equal amount of fresh RNaseIII was added, and the incubation continued for a further 30 min. RNA was isolated from the beads and from the supernatant fractions and analyzed as described above.

Gel retardation assay. Binding reactions (10 μl) were similar to those for kinase assays, except that ATP was omitted and tRNA and BSA were present at 0.1 and 1 mg/ml, respectively. The concentration of labeled dsRNA was 55 ng/ml, and the concentration of DAI (Mono S fraction), immunoaffinity chromatography-purified DAI (14), or p20 (55) was varied. After incubation for 20 min on ice, a dye-glycerol solution was added and the samples were loaded directly onto a 5% polyacrylamide gel (acrylamide: bioacrylamide, 82:1). The gel was cast in 40 mM Tris-glycine buffer and had been prerun for 1 h at 150 V. Radioactivity was detected by autoradiography for approximately 16 h.

RESULTS

Characteristics of synthetic dsRNA. Duplexed RNAs of defined sizes were made by annealing a 358-nt transcript synthesized by T7 RNA polymerase with complementary transcripts of various lengths synthesized by T3 RNA polymerase (Fig. 1A). After digestion of the RNA tails and residual single-stranded RNA, the dsRNAs were purified by electrophoresis in nondenaturing polyacrylamide gels. When analyzed in denaturing conditions (Fig. 1B), the individual strands of the dsRNA molecules were slightly heterogeneous, with chain lengths a few nucleotides longer or shorter than the input single strands as a result of the trimming process. When examined in a nondenaturing gel, however, the dsRNAs migrated as discrete bands, with mobilities similar to those of dsDNA markers (see Fig. 5A, lanes 3 to 9). As expected, the duplexes were sensitive to digestion with RNaseIII, a dsRNA-specific enzyme, but resistant to digestion by single-stranded specific nucleases except after denaturation (data not shown).

Activation and inhibition of DAI. Activation of DAI is accompanied by its autophosphorylation, converting the enzyme from a latent state to a form which can phosphorylate eIF-2α. When the synthetic duplexes were examined for their ability to catalyze autophosphorylation, we found that 23- and 34-bp dsRNAs were only slightly active, 40-bp dsRNA was partly active, and full activity was approached with 55- to 85-bp dsRNAs, which were nearly as active as the very long dsRNA (average size of >2,000 bp) isolated from reovirus virions (Fig. 2A). These results, obtained with essentially flush-ended dsRNA, agree closely with previously reported data obtained by using RNA molecules in which one strand was considerably longer than the other (44): in the earlier study, duplexes shorter than 30 bp were unable to activate DAI, and full activation was obtained with duplexes longer than 65 to 80 bp. The activation of DAI can also be monitored by phosphorylation of eIF-2, the natural substrate of this kinase. In this assay, 15-bp dsRNA was essentially inactive, 34-bp dsRNA was partially active, and 55-bp (or longer) dsRNA was fully active (Fig. 2B). Thus, the slight autophosphorylation of DAI that is catalyzed by the 34-bp duplex is sufficient to permit DAI to phosphorylate its natural substrate weakly. These results are consistent with the findings that the very short (<20 bp) imperfect duplexes found in viral RNAs such as VA RNA (28, 38, 41,

Downloaded from mcb.asm.org by on March 28, 2007

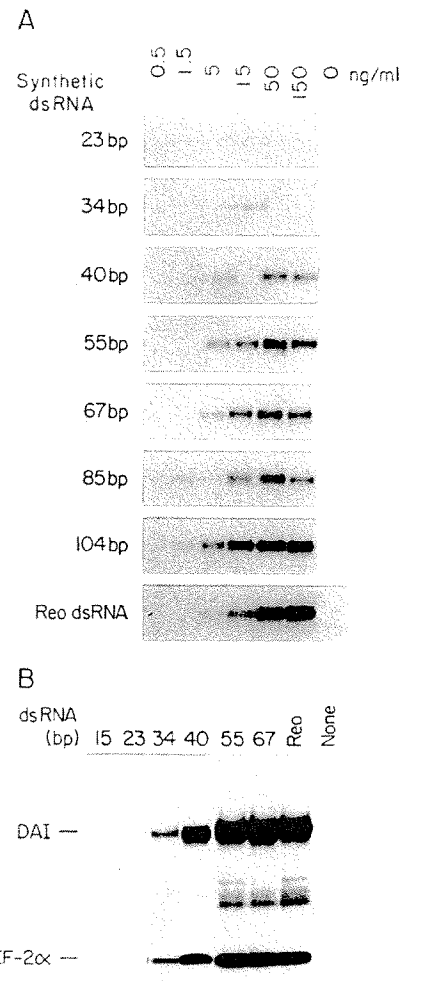


FIG. 2. Activation of DAI by short duplexes. (A) Autophosphorylation of DAI as a function of dsRNA concentration and chain length. Kinase assay reaction mixtures containing synthetic or reovirus dsRNA at the concentrations indicated were analyzed by electrophoresis in SDS-polyacrylamide gels and autoradiography. (B) Phosphorylation of eIF-2. Kinase assay reaction mixtures containing 33 ng of the synthetic dsRNAs indicated per ml or 40 ng of reovirus dsRNA per ml were supplemented with eIF-2 and were analyzed as described in part A. The positions of autophosphorylated DAI and the phosphorylated α subunit of eIF-2 are marked.

42, 48), EBER (4, 7) and TAR RNA (18) are insufficient to cause DAI activation.

From experiments with tailed duplexes, it was concluded that short dsRNAs which are incapable of activating DAI may nevertheless block the activation of DAI by longer dsRNAs (44). Figure 3A shows that at high concentration a flush-ended 23-bp dsRNA inhibited the activation of DAI by reovirus dsRNA. In this respect, short dsRNA resembles longer dsRNA molecules which block DAI activation at high concentrations, although the mechanisms might be different. Figure 3B demonstrates that long dsRNA (approximately 3,000 bp) isolated from *P. chrysogenum* activates DAI at concentrations of up to 1 μg/ml and prevents activation at 10 μg/ml, as expected from the bell-shaped activation curve.

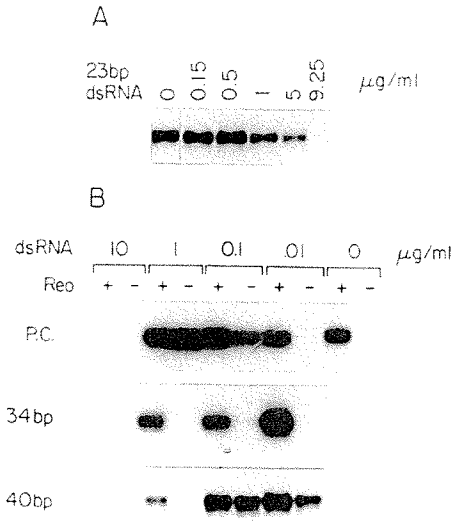


FIG. 3. Inhibition of DAI activation by high concentrations of dsRNA. (A) Inhibition of DAI autophosphorylation by 23-bp dsRNA. Kinase assay reaction mixtures contained 20 ng of reovirus dsRNA per ml and the indicated concentrations of the 23-bp dsRNA. (B) Effect of duplexes of various lengths on DAI autophosphorylation in the presence and absence of an activator dsRNA. Kinase assay reaction mixtures containing or lacking reovirus dsRNA (40 ng/ml) were supplemented with synthetic 34- or 40-bp dsRNAs or with *P. chrysogenum* dsRNA (P.C.) at the concentrations indicated. Autophosphorylation of DAI was assessed as described for Fig. 2.

Short duplexes, such as the 34-bp dsRNA (Fig. 3B), activate DAI only weakly and also become inhibitory at approximately 10 μg/ml. Similar results were obtained with bacteriophage f2 sus3 dsRNA (approximately 30 to 50 bp long [22]) and the 23-bp synthetic duplexes (data not shown), while the 40-bp dsRNA gave a response intermediate between that of the longer and shorter duplexes because of its significant ability to activate DAI (Fig. 3B). Single-stranded RNA is not inhibitory (data not shown). These results confirm that the enzyme is activated by relatively long dsRNA and can interact with short duplexes in a nonproductive way.

RNA size dependence of DAI binding. The existence of opposing effects of dsRNA on DAI activity makes it important to address directly the binding of this ligand to the protein. We examined the interactions between DAI and dsRNA, using three assays which explore different aspects of the process: nitrocellulose filter binding, binding to an affinity matrix, and gel retardation. Figure 4A and B illustrates the binding of single-stranded RNA and dsRNA as a function of chain length in the filtration assay. Duplexes with sizes of 15 and 23 bp did not bind detectably, whereas 34- and 40-bp duplexes bound weakly. Longer molecules bound with increasing efficiency, and 85- or 104-bp dsRNA bound as efficiently as 354-bp molecules. Single-stranded molecules bound only very weakly, except for the 354-nt molecule, which may be able to form a small amount of secondary structure. The binding of dsRNAs to DAI was also measured as a function of ligand concentration (Fig. 4C and D). The efficiency with which short duplexes bound to DAI did not improve at subsaturating RNA concentrations, suggesting that the affinity of the enzyme is low for molecules with sizes

Downloaded from mcb.asm.org by on March 28, 2007

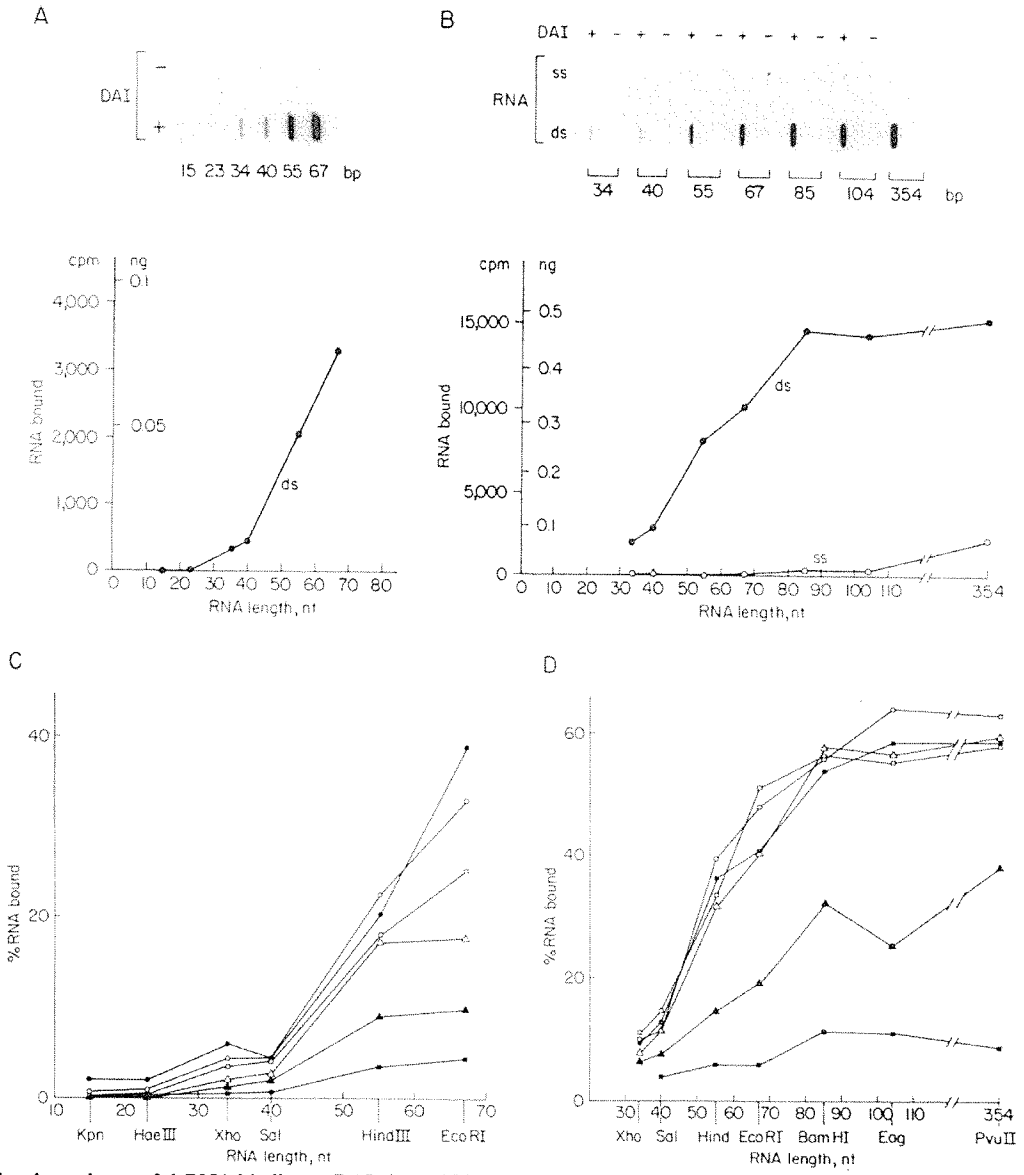


FIG. 4. Size dependence of dsRNA binding to DAI. (A and B) Nitrocellulose filter-binding assay for the binding of single-stranded RNA and dsRNAs. Autoradiograms of the membrane are shown (top panel). Quantitation of RNA binding (bottom panel) was obtained by scintillation counting or scanning of the individual filter bands and subtraction of the background value (lanes labeled −DAI) from the signal radioactivity (lanes labeled +DAI). ●, dsRNA (ds); ○, single-stranded RNA (ss). (C and D) Concentration and size dependence of dsRNA-binding efficiency. The percentage of the input dsRNA that was retained on the filter in the presence of DAI was quantified at various dsRNA concentrations of 1,000 (■), 330 (▲), 100 (△), 33 (□), 10 (○), and 3.3 (●) ng/ml, respectively.

of 40 bp or less and that the affinity increases uniformly as the chain length is increased, reaching a maximum at 85 bp. These data agree closely with the dependence of activation and inhibition on dsRNA chain length (Figs. 2 and 3) and are consistent with a model which equates activation with stable dsRNA binding. Duplexes shorter than 30 to 40 bp bind weakly and cannot activate although they inhibit activation; longer duplexes (40 to 85 bp) bind with increasing stability and their ability to activate the enzyme increases concomitantly; beyond this length, the efficiency of binding and activation remains unchanged.

Binding and protection of dsRNA. One interpretation of these observations is that the dsRNA binding site in DAI accommodates up to ~85 bp of duplex but can bind shorter duplexes less stably, down to ~30 bp. To test this interpretation we employed DAI immobilized on antibody-Sepharose beads. First, to define the minimum size of dsRNA that can bind to the enzyme, a mixture of dsRNA molecules was partially digested with RNaseIII to generate a collection of duplex molecules with a broad size distribution. This collection was allowed to bind to the immobilized DAI, and the beads were washed to remove nonspecifically adsorbed dsRNA. Figure 5A displays the DAI-bound dsRNA (lane 1) and the unbound RNA that remained in the supernatant (lane

Downloaded from mcb.asm.org by on March 28, 2007

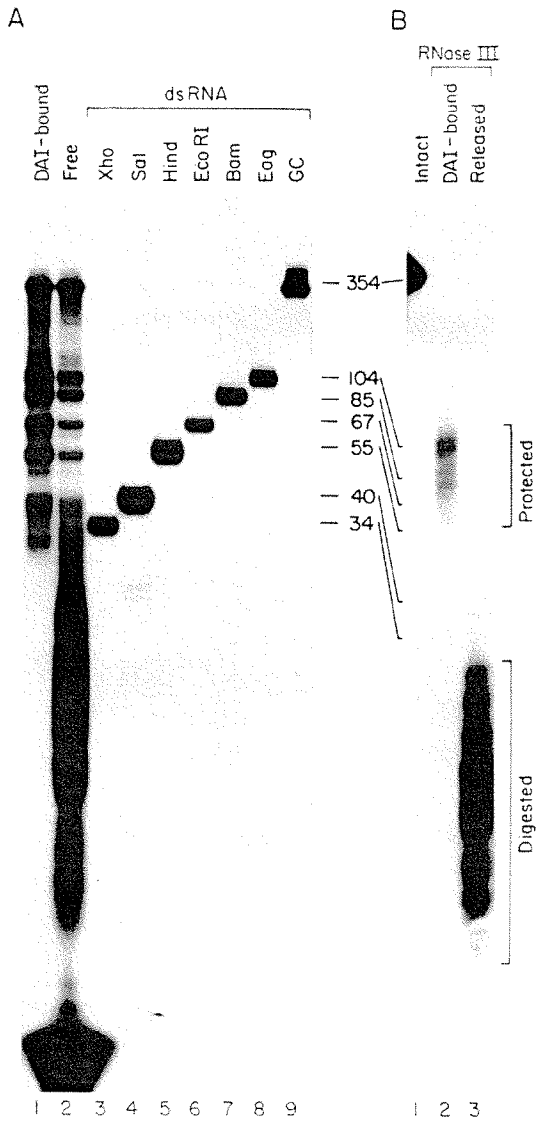


FIG. 5. Binding and protection of dsRNA fragments by DAI. (A) Binding of randomly sized dsRNA fragments. A mixture of fragments (approximately 10 to 350 bp) was incubated with DAI attached to Sepharose beads. Unbound RNA was separated from the beads by centrifugation and washing. Equal fractions of the DAI-bound (lane 1) and unbound (lane 2) dsRNA were resolved in a nondenaturing polyacrylamide gel and detected by autoradiography. Markers included discrete dsRNA fragments with sizes of 34 to 104 bp (lanes 3 to 8), denoted by the restriction site designation used in their synthesis (Fig. 1A), and 354 bp (lane 9), denoted GC for the plasmid used in its synthesis (pGEM.GC). (B) Protection of DAI-bound dsRNA from digestion by RNaseIII. Discrete 354-bp dsRNA (lane 1) was bound to DAI attached to Sepharose beads, and unbound RNA was removed. The beads were exhaustively incubated with RNaseIII, and the released RNA fragments were collected. Equal fractions of the released RNA (lane 3) and the RNA that remained associated with the beads (lane 2) were resolved as described for panel A.

2). Comparison with dsRNA markers (lanes 3 to 9) and with an RNA sequence ladder (data not shown), indicated that the cutoff for binding was at approximately 28 bp, in good agreement with results obtained in the nitrocellulose filter-binding assay. Moreover, visual inspection of the autoradiogram suggested that dsRNA with a size of 28 to 40 bp bound less efficiently than longer duplexes.

Next, we conducted a protection experiment to determine the length of dsRNA that is shielded by DAI from nuclease attack. Intact 354-bp dsRNA was bound to DAI immobilized on antibody-Sepharose beads, and the excess unbound dsRNA was removed. The bound dsRNA was digested by incubating the beads with RNaseIII to trim off regions of duplex that were not protected by DAI. Figure 5B, lane 3, shows that the released dsRNA had been reduced to fragments of approximately 10 to 20 bp as expected (57), whereas the bulk of the DAI-associated material (lane 2) ranged in size from approximately 60 to 120 bp, with a substantial concentration in the longer-size class (approximately 100 to 120 bp). Assuming that RNaseIII leaves 15 bp of dsRNA protruding on each side, we deduce that DAI associates with 30 to 90 bp of dsRNA. Taking 110 bp as the modal length of the protected fragments, it appears that about 80 bp of duplex interact directly with the enzyme, roughly the length of dsRNA that gives maximal binding in the nitrocellulose filter assay. The length of the protected fragment was not altered at relatively high concentrations of dsRNA (up to 1 μg/ml; data not shown), conditions which would be expected to disfavor oligomerization of DAI on the dsRNA. These findings support the view that the dsRNA site extends for ~80 bp and that shorter molecules bind with lesser affinity, provided that they are at least 28 bp long.

DAI-dsRNA complexes. To characterize the interactions more directly, we examined complexes formed between DAI and dsRNAs in a gel retardation assay (Fig. 6A). No complexes were observed with 15- or 23-bp duplexes (data not shown), but longer dsRNAs formed complexes with increasing efficiency. Four series of complexes were distinguishable (bands I to IV). Their relative abundance was principally a function of RNA chain length, with a lesser dependence on DAI concentration. On the basis of their behavior, the complexes seem to fall into two families. One family, containing the more slowly moving bands I and II, forms preferentially with duplexes of less than optimal length (34 to 67 bp) in binding and activation assays. The second family, containing the faster moving bands III and IV, forms preferentially with longer duplexes (≥85 bp), which are fully active in binding and activating DAI.

Band I was the most prominent complex with 55- and 67-bp duplexes but was barely detectable with longer or shorter duplexes. It was formed at low DAI concentrations and seemed to be converted to band II at elevated DAI concentrations. Indeed, regardless of chain length, complex II was seen only at high DAI concentrations. With 34- and 40-bp duplexes, the only detectable complexes appeared to migrate in band II. Formation of complex II increased as the chain length was extended, reaching maximal levels with the 67-bp duplex and declining as the chain length was extended further, to 85 and 104 bp.

The most abundant complexes, formed with 85- and 104-bp duplexes, migrated in band III. This band was also visible with 67-bp and perhaps 55-bp dsRNAs. As chain length increased, complex III was formed at progressively lower concentrations of DAI. It also seemed to decrease slightly at high DAI concentrations. Band IV displayed a pattern similar to that of band III but was always less

Downloaded from mcb.asm.org by on March 28, 2007

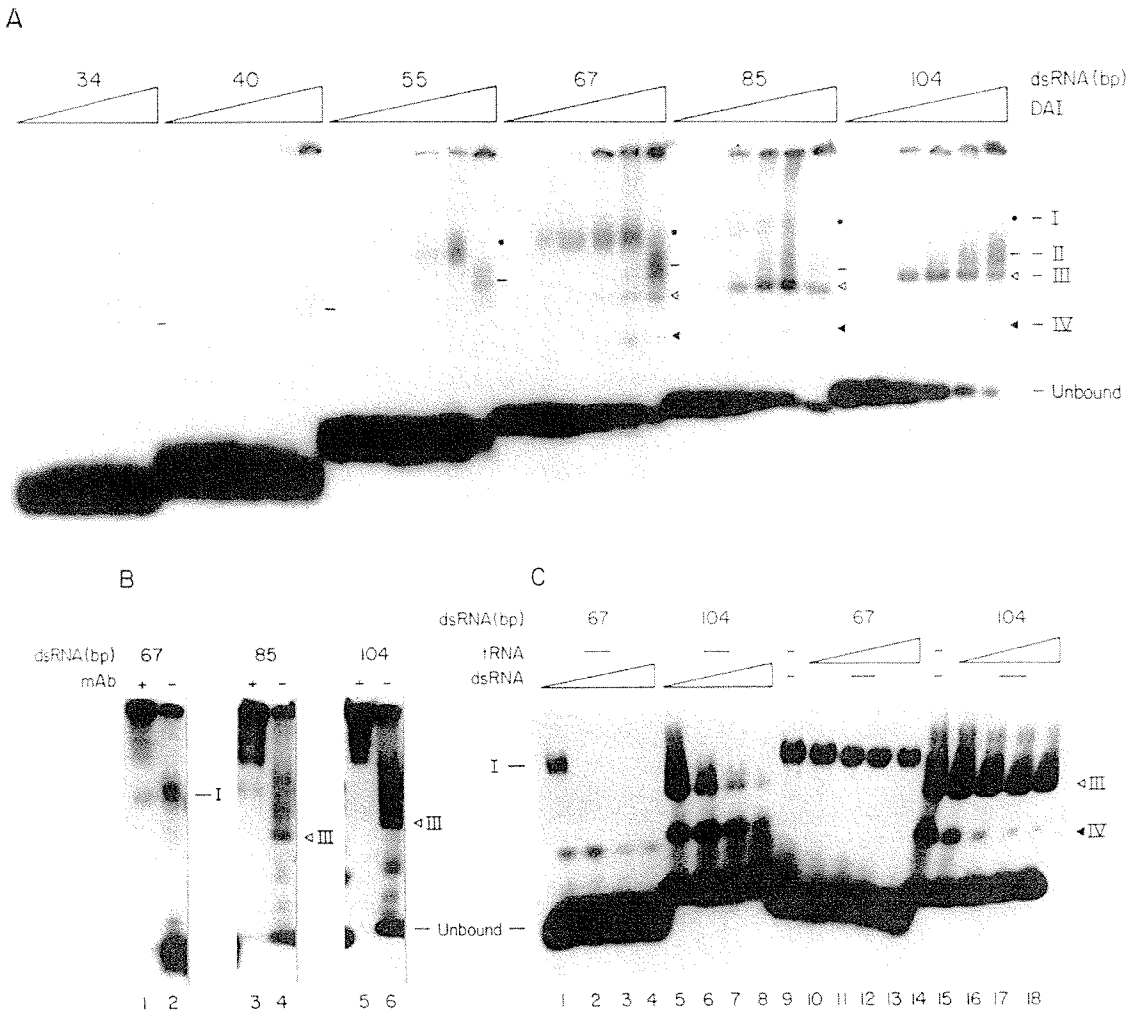


FIG. 6. Gel-shift analysis with DAI. (A) Dependence on dsRNA size and DAI concentration. dsRNAs with sizes of 34 to 104 bp were incubated with various amounts of DAI (0, 0.05, 0.1, 0.25, 0.5, and 1.0 μl from left to right, symbolized by the wedges) purified to the Mono S stage. The resultant complexes were separated by electrophoresis in nondenaturing conditions and detected by autoradiography. The positions of complexes I (●), II (—), III (◁), and IV (◀) and of the free dsRNA fragments are marked in each panel. (B) Shifts with essentially homogeneous DAI and antibody supershift. dsRNA with a size of 67 bp (lanes 1 and 2), 85 bp (lanes 3 and 4), or 104 bp (lanes 5 and 6) was incubated with 1 μl of DAI purified by immunoaffinity chromatography in the presence (lanes 1, 3, and 5) or absence (lanes 2, 4, and 6) of 1 μl of monoclonal antibody to DAI. (C) Competition assays. Standard reaction mixtures (lanes 9 and 14) contained DAI (Mono S fraction) and 100 μg of tRNA per ml; *P. chrysogenum* dsRNA (0.25, 0.5, 0.75, and 1 μg/ml; lanes 1 to 4 and 5 to 8) or additional calf liver tRNA (100, 200, 300 and 400 μg/ml; lanes 10 to 13 and 15 to 18) was added as indicated. ^{32}P-labeled dsRNA (67 bp [lanes 1 to 4 and 9 to 13] or 104 bp [lanes 5 to 8 and 14 to 18]) was present at 55 ng/ml. The wedges symbolize increasing concentrations from left to right; — indicates absence of the RNA.

abundant. Both of these bands correlate well with full enzyme activity.

Very similar patterns of bands were formed with an essentially homogeneous preparation of DAI purified by immunoaffinity chromatography (Fig. 6B), and the gel-shift activity cosedimented with kinase activity through a glycerol gradient (29) (data not shown). Furthermore, as seen in Fig. 6B, all of the complexes were "supershifted" to forms with slower mobility by addition of monoclonal antibody directed against DAI. The antibody did not produce a gel shift on its own (i.e., in the absence of DAI), but it appeared to stabilize DAI-dsRNA complexes so that less probe remained in free form. These experiments verified that bands I to IV all contain DAI. Competition experiments demonstrated that the most prominent complexes, band I with 67-bp dsRNA and band III with 104-bp dsRNA, were resistant to the presence of excess tRNA but were sensitive to unlabeled *P. chrysogenum* dsRNA competitor (Fig. 6C). The complexes formed with the 104-bp dsRNA were more resistant to competition than those formed with 67-bp dsRNA, as expected from the higher affinity of DAI for longer duplexes (Fig. 4). For some of the minor bands (e.g., band IV), competition was more effective with tRNA at very high concentrations than with dsRNA at moderate concentrations

Downloaded from mcb.asm.org by on March 28, 2007

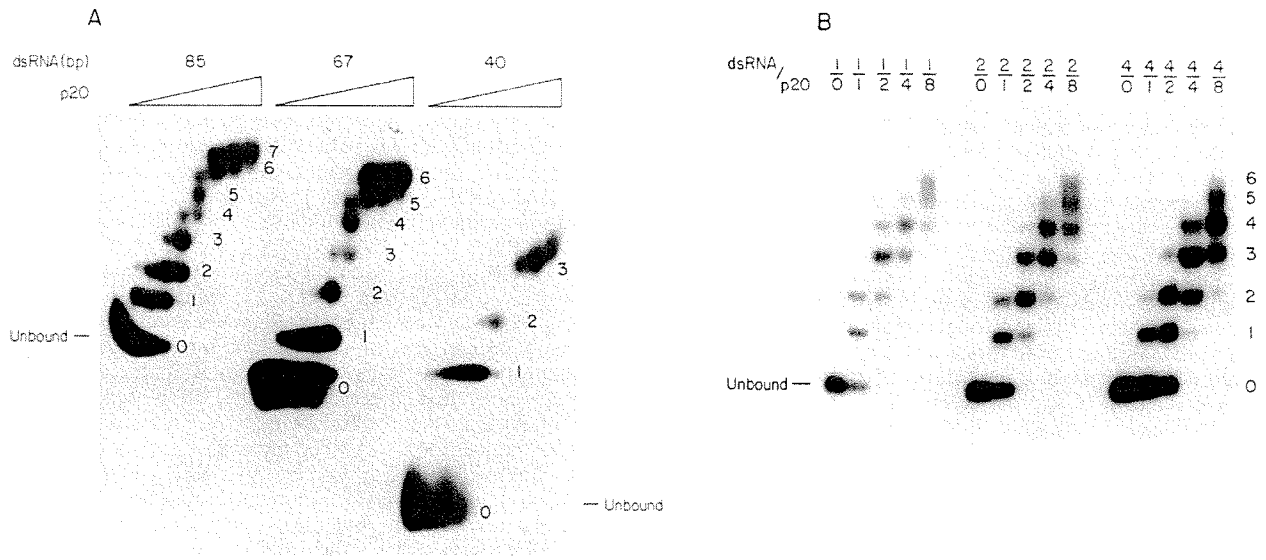


FIG. 7. Gel shift analysis with the p20 polypeptide. (A) Dependence on dsRNA size. Reaction mixtures contained dsRNA of 40, 67, or 85 bp as indicated and 0, 0.032, 0.063, 0.125, 0.25, 0.5, 2.5, 5, or 10 μg/ml of p20 (increasing from left to right as symbolized by the wedges). The number of p20 units in each band is indicated on the right. (B) Dependence on dsRNA and p20 concentration. Assay mixtures contained 85-bp dsRNA at a concentration of 55, 110, or 220 μg/ml and p20 at a concentration of 0, 63, 125, 250, or 500 ng/ml. Concentrations increase from left to right. The relative concentrations of these two components are indicated above the autoradiogram, and the number of p20 units in each band is marked on the side.

for reasons that are presently unclear. Neither single-stranded DNA nor dsDNA was an effective competitor (data not shown). These results demonstrate that the most prominent complexes are dsRNA and DAI specific and that there is a rather abrupt change in electrophoretic mobility when the dsRNA reaches the size for optimal binding and activation (approximately 80 bp).

Minimal binding site. The ability of short duplexes (<30 bp) to inhibit DAI activation implies that they interact with the enzyme, even though their binding is not sufficiently stable to be detected by the assays used to this point. To determine the minimal length of duplex that can interact with DAI, we employed the close-packing method for estimating the number of protein molecules that can bind to duplexes with known sizes. For this purpose, we used a truncated version of DAI, p20, comprising the N-terminal 184 amino acids which we and others have determined to contain the RNA-binding domain of the protein (16, 25, 39, 50). The numbers of p20 molecules binding to a given dsRNA were estimated from gel-shift assays conducted at increasing concentrations of the protein. As seen in Fig. 7A, a series of complexes was formed, reaching a maximum at the highest protein levels. The number of complexes increased with increasing dsRNA chain length as follows: 40 bp, three bands; 67 bp, six bands; 85 bp, seven bands. Assuming that each shifted band corresponds to the binding of a p20 molecule, these data imply that the minimum binding site is about 11 bp, equivalent to a single turn of A-form RNA helix.

The appearance of a ladder of bands with p20 suggested that there may be differences between the binding of this fragment and intact DAI to dsRNA. Further experiments showed that the formation of the p20 complexes is specific in that tRNA does not compete (data not shown). To rule out the possibility that the p20 banding patterns are due to a concentration-dependent protein oligomerization that is independent of dsRNA, we conducted band-shift assays at increasing dsRNA concentrations. Figure 7B shows that more p20 protein is required to achieve a given gel shift at higher concentrations of dsRNA (compare, for example, the amount of p20 needed to complex all of the dsRNA in the reactions). This indicates that oligomerization depends on the presence of dsRNA and is due to the formation of a series of protein-RNA complexes rather than to preformed protein-protein aggregates. The banding pattern was also influenced by the absolute concentration of p20 and dsRNA, however, as can be seen by comparing lanes with equal ratios of dsRNA to p20 (such as 1/1, 2/2, and 4/4). The shift to larger complexes at higher concentrations could merely reflect the concentration dependence of the reaction according to the law of mass action, or it could imply that p20 complexes are stabilized by protein-protein interactions when p20 monomers are bound adjacently on dsRNA. In the latter case, stabilizing protein-protein interactions would provide an explanation for the apparent paradox that DAI binds efficiently only to duplexes of longer than 30 bp but can bind to as little as a single helical turn of dsRNA.

DISCUSSION

Although the existence of DAI has been known for many years and its activation by a variety of polynucleotides has been studied intensively, an understanding of the enzyme's regulation has remained elusive. The kinase is activated by autophosphorylation in the presence of dsRNA. This response exhibits a number of unusual features: first, activation is prevented by high concentrations of dsRNAs which activate the kinase at low concentrations; second, short RNA duplexes fail to activate DAI at any concentration but prevent activation at elevated concentrations; third, highly structured single-stranded RNAs of viral origin also fail to activate DAI but can block activation by authentic, long

Downloaded from mcb.asm.org by on March 28, 2007

dsRNA. To illuminate the interactions between dsRNA and DAI, we generated a series of short RNA duplexes and studied directly their binding to the enzyme as well as their effects on its activity. The results correlate activation with the formation of stable complexes with a characteristic electrophoretic mobility and suggest a model that is compatible with the emerging understanding of DAI structure.

Our results are most consistent with the view that DAI possesses a single effective site for dsRNA, capable of accommodating approximately 80 bp of duplex. Two observations support this conclusion most strongly. First, as the length of the dsRNA ligand is increased, maximal binding is attained at this size and longer molecules bind no more efficiently, and second, the kinase protects this length of duplex from digestion by nuclease. Shorter duplexes, down to a lower limit of approximately 30 bp, bind with steadily decreasing efficiency while duplexes with lengths of less than 30 bp are unable to form a stable complex with DAI under normal conditions. Nonetheless, since such very short duplexes block the activation of DAI, we assume that at high concentrations they form transient interactions which prevent DAI activation. Likewise, other polynucleotides, such as RNA-DNA hybrids and partially methylated dsRNA duplexes that fail to activate the kinase, as well as long dsRNAs that can activate DAI, share this property of inhibiting kinase activation at high concentrations. The nature of these inhibitory interactions is unclear, and it remains to be seen whether viral effectors such as VA RNA, EBER, and TAR RNA function in the same way as short duplexes or whether they interact in a distinct fashion to block DAI activation. Preliminary data indicate that the sites for VA RNA and dsRNA are overlapping but perhaps not congruent (16).

How do these functional observations relate to the structure of the enzyme? DAI possesses two RNA-binding elements in its N-terminal domain (12, 16, 25, 39, 50). Each element contains an RNA-binding motif which is rich in basic amino acids and is predicted to form an α-helical structure (16, 39). Both elements are required for efficient binding of RNA, and they appear to cooperate to form a single bivalent site which optimally extends over approximately 80 bp of duplex. Since the RNA binding domain of DAI, expressed as the p20 protein, is able to interact with as little as 11 bp of dsRNA, we speculate that each element interacts with a single helical turn and that optimal binding occurs when these two turns are separated by about five intervening helical turns. In this complex, the entire span of approximately 80 bp is protected by DAI against attack by the dsRNA-specific nuclease RNaseIII. With this model, interactions with shorter dsRNA molecules entail increasing strain on the enzyme, accompanied by decreasing affinity, such that it becomes impossible for both elements to bind when there is less than one intervening helical turn (at approximately 33 bp). Evidently, monovalent complexes can also be formed at high ratios of enzyme to RNA as in the p20 gel-shift experiments: these complexes presumably involve only the stronger RNA binding region (region 1 [16]) and allow the protein to pack onto the RNA to a density of one molecule per helical turn.

According to this model, activation of the enzyme requires bivalent dsRNA binding which becomes detectable at approximately 30 bp and is most stable when the duplex is at least 80 bp long. Correlating with the formation of the most stable complexes is a shift in their mobility in the gel retardation assay. The predominant complex formed with dsRNA with a length of ≥85 bp is band III, which moves faster than the predominant complex formed with shorter dsRNA (band I). We considered the possibility that longer duplexes might be able to bind more DAI molecules than shorter duplexes, but because DAI is a basic protein (pI 8.6), it is unlikely that the acceleration in gel mobility that occurs between 67 and 85 bp with the shift from complex I to complex III is due to the binding of a second DAI molecule to a DAI-dsRNA complex. Therefore, we argue that the faster migration is probably due to a conformational change in the dsRNA or the DAI-dsRNA complex which leads to compaction and increased electrophoretic mobility. Compaction could result from relief of the distortion in DAI that occurs when the two binding elements can interact with optimally spaced sites on dsRNA. Alternatively, it could be accomplished if the RNA were bent or wrapped around the enzyme once it had filled the entire site. If this explanation is correct, it seems that the duplex must be continuous since an elevated concentration of 40-bp molecules does not have the same effect on binding or activation as an 80-bp duplex. The minor complexes, II and IV, seem to be related to complexes I and III, respectively, but display increased sensitivity to competition with tRNA. They are unlikely to represent the addition of a second molecule of DAI to a DAI-dsRNA complex because of the large retardation effect that this would be expected to have on electrophoretic mobility and there are few clues as to their structure or significance at present.

The proposal that DAI contains a single bipartite RNA-binding site provides an alternative to the two previous models for DAI activation, neither of which is readily compatible with the results presented here. The gel-shift data could be interpreted in terms of the model in which DAI possesses two distinct sites for dsRNA binding (15, 37), a high-affinity site for activation and a low-affinity inhibitory site, if it were supposed that duplexes with sizes of ≤67 bp bind at the inhibitory site whereas longer duplexes bind at the activating site. The resultant complexes could have significantly different mobilities. However, the apparent affinities for these duplexes are not greatly different in the binding assays shown here, so the postulate that the activation site is of much higher affinity than the inhibitory site is not satisfied. Moreover, 40- to 67-bp duplexes have significant ability to activate the enzyme. The data could also be interpreted in terms of the model that DAI is activated when two molecules bind to a single molecule of dsRNA (37, 49). On this basis, 80 bp would be the length of duplex required to span the RNA-binding sites of two DAI monomers. Each monomer would interact with 30 to 40 bp of dsRNA, and the complex would be stabilized by cooperative interactions between the protein molecules. The monotonous increase in binding efficiency with a chain length between 30 and 85 bp argues against this model, as does our failure to obtain protection of shorter RNA fragments at high ratios of DAI to dsRNA. Also, with this model, it is difficult to explain the multiplicity of complexes observed in band-shift experiments: in particular, longer duplexes or higher DAI concentrations would be expected to give rise to slower complexes, contrary to observations.

In summary, the data presented here suggest that DAI interacts with as little as 11 bp (one helical turn) of dsRNA, but activation is associated with the formation of a stable DAI-dsRNA complex. The formation of such a complex requires at least 30 bp of duplex (about three turns) and probably takes place when both of the enzyme's RNA-binding motifs are engaged with the ligand. Complex formation is optimal with dsRNA containing at least 80 bp (seven

Downloaded from mcb.asm.org by on March 28, 2007

to eight turns) and is apparently accompanied by a conformational change in the complex. We speculate that the bivalent interaction with dsRNA or the conformational change itself is critical for enzyme autophosphorylation and activation. With this model, short dsRNAs would be expected to block activation because they can interact with only one RNA-binding motif. Specific inhibitors of DAI activation, such as VA RNA, may also bind to one motif, or alternatively they may bind to both motifs but in such a way as to interfere with the conformation of the enzyme. Similarly, long dsRNAs that activate the enzyme at low concentrations might block activation at high concentrations because the two binding motifs form complexes with separate RNA duplexes, thereby precluding the requisite conformational change. While this is a satisfying explanation, there are alternatives which also fit the available facts. For example, since autophosphorylation appears to be intermolecular (29, 37), it is also possible that DAI serves as a phosphate acceptor only when it is not bound to dsRNA, a situation which would obtain at low or moderate concentrations of dsRNA. Consistent with this view, a truncated form of DAI that lacks the RNA binding site can still be phosphorylated by intact DAI (unpublished data). It is clear that further investigation will be required to establish the nature of the coupling between dsRNA binding and kinase activation: such studies are in progress.

ACKNOWLEDGMENTS

We thank P. Clarke, Y. Ma, K. Mellits, and T. Pe'ery for helpful discussions, A. Shatkin for reovirus dsRNA, A. Hovanessian for monoclonal antibody and antibody-Sepharose, and H. Robertson for *P. chrysogenum* dsRNA, f2 sus3 dsRNA, and RNase III.

Christian Schmedt was a member of the University of Konstanz/SUNY at Stony Brook exchange program. This work was supported by NIH grant CA13106.

REFERENCES

1. **Baglioni, C., and P. A. Maroney.** 1981. Inhibition of double-stranded ribonucleic acid activated protein kinase and 2',5'-oligo(adenylic acid) polymerase by ethidium bromide. Biochemistry **20**:758–762.
2. **Baglioni, C., M. A. Minks, and E. DeClercq.** 1981. Structural requirements of polynucleotides for the activation of $(2'\text{-}5')A_n$ polymerase and protein kinase. Nucleic Acids Res. **9**:4939–4950.
3. **Berry, M. J., G. S. Knutson, S. R. Laskey, S. M. Munemitsu, and C. E. Samuel.** 1985. Purification and substrate specificities of the double-stranded RNA-dependent protein kinase from untreated and interferon-treated mouse fibroblasts. J. Biol. Chem. **260**:11240–11247.
4. **Bhat, R. A., and B. Thimmappaya.** 1985. Construction and analysis of additional adenovirus substitution mutants confirm the complementation of VAI RNA function by two small RNAs encoded by Epstein-Barr virus. J. Virol. **56**:750–756.
5. **Bischoff, J. R., and C. E. Samuel.** 1989. Mechanism of interferon action. Activation of the human P1/eIF-2α protein kinase by individual reovirus s-class mRNAs: s1 mRNA is a potent activator relative to s4 mRNA. Virology **172**:106–115.
6. **Chong, K. L., L. Feng, K. Schappert, E. Meurs, T. F. Donahue, J. D. Friesen, A. G. Hovenessian, and B. R. G. Williams.** 1992. Human p68 kinase exhibits growth suppression in yeast and homology to the translational regulator *GCN2*. EMBO J. **11**: 1553–1562.
7. **Clarke, P. A., N. A. Sharp, and M. J. Clemens.** 1990. Translational control by the Epstein-Barr virus small RNA EBER-1. Reversal of the double-stranded RNA-induced inhibition of protein synthesis in reticulocyte lysates. Eur. J. Biochem. **193**:635–641.
8. **deBenedetti, A., and C. Baglioni.** 1984. Inhibition of mRNA binding to ribosomes by localized activation of dsRNA-dependent protein kinase. Nature (London) **311**:79–81.
9. **Dever, T. E., L. Feng, R. C. Wek, A. M. Cigan, T. F. Donahue, and A. G. Hinnebusch.** 1992. Phosphorylation of initiation factor 2α by protein kinase GCN2 mediates gene-specific translational control of *GCN4* in yeast. Cell **68**:585–596.
10. **Dubois, M.-F., J. Galabru, P. Lebon, B. Safer, and A. G. Hovanessian.** 1989. Reduced activity of the interferon-induced double-stranded RNA-dependent protein kinase during a heat shock stress. J. Biol. Chem. **264**:12165–12171.
11. **Farrell, P. J., K. Balkow, T. Hunt, and R. J. Jackson.** 1977. Phosphorylation of initiation factor eIF-2 and the control of reticulocyte protein synthesis. Cell **11**:187–200.
12. **Feng, G.-S., K. Chong, A. Kumar, and B. R. G. Williams.** 1992. Identification of double-stranded RNA-binding domains in the interferon-induced double-stranded RNA-activated p68 kinase. Proc. Natl. Acad. Sci. USA **89**:5447–5451.
13. **Galabru, J., and A. Hovanessian.** 1985. Two interferon-induced proteins are involved in the protein kinase complex dependent on double-stranded RNA. Cell **43**:685–694.
14. **Galabru, J., and A. Hovanessian.** 1987. Autophosphorylation of the protein kinase dependent on double-stranded RNA. J. Biol. Chem. **262**:15538–15544.
15. **Galabru, J., M. G. Katze, N. Robert, and A. G. Hovanessian.** 1989. The binding of double-stranded RNA and adenovirus VAI RNA to the interferon-induced protein kinase. Eur. J. Biochem. **178**:581–589.
16. **Green, S. R., and M. B. Mathews.** Submitted for publication.
17. **Grodberg, J., and J. J. Dunn.** 1988. *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. J. Bacteriol. **170**:1245–1253.
18. **Gunnery, S., A. P. Rice, H. D. Robertson, and M. B. Mathews.** 1990. Tat-responsive region RNA of human immunodeficiency virus 1 can prevent activation of the double-stranded-RNA-activated protein kinase. Proc. Natl. Acad. Sci. USA **87**:8687–8691.
19. **Hershey, J. W. B.** 1989. Protein phosphorylation controls translation rates. J. Biol. Chem. **264**:20823–20826.
20. **Hershey, J. W. B.** 1991. Translational control in mammalian cells. Annu. Rev. Biochem. **60**:717–755.
21. **Hovanessian, A. G.** 1989. The double stranded RNA-activated protein kinase induced by interferon: dsRNA-PK. J. Interferon Res. **9**:641–647.
22. **Hunter, T., T. Hunt, R. J. Jackson, and H. D. Robertson.** 1975. The characteristics of inhibition of protein synthesis by double-stranded ribonucleic acid in reticulocyte lysates. J. Biol. Chem. **250**:409–417.
23. **Judware, R., and R. Petryshyn.** 1991. Partial characterization of a cellular factor that regulates the double-stranded RNA-dependent eIF-2α kinase in 3T3-F442A fibroblasts. Mol. Cell. Biol. **11**:3259–3267.
24. **Katze, M. G., B. M. Detjen, B. Safer, and R. M. Krug.** 1986. Translational control by influenza virus: suppression of the kinase that phosphorylates the alpha subunit of initiation factor eIF-2 and selective translation of influenza viral messenger RNAs. Mol. Cell. Biol. **6**:1741–1750.
25. **Katze, M. G., M. Wambach, M.-L. Wong, M. Garfinkel, E. Meurs, K. Chong, B. R. G. Williams, A. G. Hovanessian, and G. N. Barber.** 1991. Functional expression and RNA binding analysis of the interferon-induced, double-stranded RNA-activated, 68,000-M_r protein kinase in a cell-free system. Mol. Cell. Biol. **11**:5497–5505.
26. **Kaufman, R. J., and P. Murtha.** 1987. Translational control mediated by eukaryotic initiation factor-2 is restricted to specific mRNAs in transfected cells. Mol. Cell. Biol. **7**:1568–1571.
27. **Kimchi, A., A. Zilberstein, A. Schmidt, L. Shulman, and M. Revel.** 1979. The interferon-induced protein kinase PK-i from mouse L cells. J. Biol. Chem. **254**:9846–9853.
28. **Kitajewski, J., R. J. Schneider, B. Safer, S. M. Munemitsu, C. E. Samuel, B. Thimmappaya, and T. Shenk.** 1986. Adenovirus VAI RNA antagonizes the antiviral action of the interferon by preventing activation of the interferon-induced eIF-2α kinase. Cell **45**:195–200.

Downloaded from mcb.asm.org by on March 28, 2007

29. **Kostura, M., and M. B. Mathews.** 1989. Purification and activation of the double-stranded RNA-dependent eIF-2 kinase DAI. Mol. Cell. Biol. **9**:1576–1586.
30. **Lasky, S. R., B. L. Jacobs, and C. E. Samuel.** 1982. Characterization of the sites of phosphorylation in the interferon-induced phosphoprotein P1 from mouse fibroblasts: evidence for two forms of P1. J. Biol. Chem. **257**:11087–11093.
31. **Laurent, A. G., B. Krust, J. Galabru, J. Svab, and A. G. Hovanessian.** 1985. Monoclonal antibodies to an interferon-induced M_r 68,000 protein and their use for the detection of double-stranded RNA-dependent protein kinase in human cells. Proc. Natl. Acad. Sci. USA **82**:4341–4345.
32. **Lebleu, B., G. C. Sen, S. Shaila, B. Cabrer, and P. Lengyel.** 1976. Interferon, double-stranded RNA, and protein phosphorylation. Proc. Natl. Acad. Sci. USA **73**:3107–3111.
33. **Lenz, J. R., and C. Baglioni.** 1978. Inhibition of protein synthesis by double-stranded RNA and phosphorylation of initiation factor, eIF-2. J. Biol. Chem. **253**:4219–4223.
34. **Levin, D., and I. M. London.** 1978. Regulation of protein synthesis: activation by double-stranded RNA of a protein kinase that phosphorylates eukaryotic initiation factor 2. Proc. Natl. Acad. Sci. USA **75**:1121–1125.
35. **Levin, D. H., R. Petryshyn, and I. M. London.** 1980. Characterization of double-stranded-RNA-activated kinase that phosphorylates α subunit of eukaryotic initiation factor 2 (eIF-2α) in reticulocyte lysates. Proc. Natl. Acad. Sci. USA **77**:832–836.
36. **Mathews, M. B.** 1991. Control of translation in adenovirus infected cells. Enzyme **44**:250–264.
37. **Mathews, M. B., S. Gunnery, L. Manche, K. H. Mellits, and T. Pe'ery.** 1990. Control of protein synthesis by RNA regulators. NATO ASI (Adv Sci Inst) Ser H Cell Biol. **49**:377–388.
38. **Mathews, M. B., and T. Shenk.** 1991. Adenovirus virus-associated RNA and translational control. J. Virol. **65**:5657–5662.
39. **McCormack, S. J., D. C. Thomis, and C. E. Samuel.** 1992. Mechanism of interferon action: identification of a RNA binding domain within the N-terminal region of the human RNA-dependent P1/eIF-2α protein kinase. Virology **188**:47–56.
40. **Mellits, K. H., M. Kostura, and M. B. Mathews.** 1990. Interaction of adenovirus VA RNA_I with the protein kinase DAI: non-equivalence of binding and function. Cell **61**:843–852.
41. **Mellits, K. H., and M. B. Mathews.** 1988. Effects of mutations in stem and loop regions on the structure and function of adenovirus VA RNA_I. EMBO J. **7**:2849–2859.
42. **Mellits, K. H., T. Pe'ery, L. Manche, H. D. Robertson, and M. B. Mathews.** 1990. Removal of double-stranded contaminants from RNA transcripts: synthesis of adenovirus VA RNA_I from a T7 vector. Nucleic Acids Res. **18**:5401–5406.
43. **Meurs, E., K. Chong, J. Galabru, N. S. B. Thomas, I. M. Kerr, B. R. G. Williams, and A. G. Hovanessian.** 1990. Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. Cell **62**: 379–390.
44. **Minks, M. A., D. K. West, S. Benvin, and C. Baglioni.** 1979. Structural requirements of double-stranded RNA for the activation of 2′,5′-oligo(A) polymerase and protein kinase of interferon-treated HeLa cells. J. Biol. Chem. **254**:10180–10183.
45. **Minks, M. A., D. K. West, S. Benvin, J. J. Greene, P. O. P. Ts'o, and C. Baglioni.** 1980. Activation of 2′,5′-oligo(A) polymerase and protein kinase of interferon-treated HeLa cells by 2′-O-methylated poly (inosinic acid) · poly (cytidylic acid). J. Biol. Chem. **255**:6403–6407.
46. **Ochoa, S.** 1983. Regulation of protein synthesis initiation in eukaryotes. Arch. Biochem. Biophys. **223**:325–349.
47. **O'Malley, R. P., R. F. Duncan, J. W. B. Hershey, and M. B. Mathews.** 1989. Modification of protein synthesis initiation factors and shut-off of host protein synthesis in adenovirus-infected cells. Virology **168**:112–118.
48. **O'Malley, R. P., T. M. Mariano, J. Siekierka, and M. B. Mathews.** 1986. A mechanism for the control of protein synthesis by adenovirus VA RNA_I. Cell **44**:391–400.
49. **O'Malley, R. P., T. M. Mariano, J. Siekierka, W. C. Merrick, P. A. Reichel, and M. B. Mathews.** 1986. The control of protein synthesis by adenovirus VA RNA. Cancer Cells **4**:291–301.
50. **Patel, R. C., and G. C. Sen.** 1992. Identification of the double-stranded RNA-binding domain of the human interferon-inducible protein kinase. J. Biol. Chem. **267**:7671–7676.
51. **Pestka, S., J. A. Langer, K. C. Zoon, and C. E. Samuel.** 1987. Interferons and their actions. Annu. Rev. Biochem. **56**:727–777.
52. **Petryshyn, R., J.-J. Chen, and I. M. London.** 1984. Growth-related expression of a double-stranded RNA-dependent protein kinase in 3T3 cells. J. Biol. Chem. **259**:14736–14742.
53. **Rice, A. R., M. Kostura, and M. B. Mathews.** 1989. Identification of a 90 KD polypeptide which associates with adenovirus VA RNA_I and is phosphorylated by the double-stranded RNA dependent protein kinase. J. Biol. Chem. **264**:20632–20637.
54. **Samuel, C. E.** 1991. Antiviral actions of interferon: interferon-regulated cellular proteins and their surprisingly selective antiviral activities. Virology **183**:1–11.
55. **Schmedt, C., S. R. Green, and M. B. Mathews.** Unpublished data.
56. **Schneider, R. J., and T. Shenk.** 1987. Impact of virus infection on host cell protein synthesis. Annu. Rev. Biochem. **56**:317–332.
57. **Schweitz, H., and J. P. Ebel.** 1971. A study of the mechanism of action of E. coli ribonuclease III. Biochimie **53**:585–593.
58. **Sen, G. C., H. Taira, and P. Lengyel.** 1978. Interferon, double-stranded RNA, and protein phosphorylation. J. Biol. Chem. **253**:5915–5921.
59. **Sonenberg, N.** 1990. Measures and countermeasures in the modulation of initiation factor activities by viruses. New Biol. **2**:402–409.
60. **Svensson, C., and G. Akusjärvi.** 1985. Adenovirus VA RNA_I mediates a translational stimulation which is not restricted to the viral mRNAs. EMBO J. **4**:957–964.
61. **Tiwari, R. K., J. Kusari, R. Kumar, and G. C. Sen.** 1988. Gene induction by interferons and double-stranded RNA: selective inhibition by 2-aminopurine. Mol. Cell. Biol. **8**:4289–4294.
62. **Torrence, P. F., M. I. Johnston, D. A. Epstein, H. Jacobsen, and R. M. Friedman.** 1981. Activation of human and mouse 2-5A synthetases and mouse protein P_1 kinase by nucleic acids. FEBS Lett. **130**:291–296.
63. **Zilberstein, A., A. Kimchi, A. Schmidt, and M. Revel.** 1978. Isolation of two interferon-induced translational inhibitors: a protein kinase and an oligo-isoadenylate synthetase. Proc. Natl. Acad. Sci. USA **75**:4734–4738.
64. **Zinn, K., A. Keller, L.-A. Whitemore, and T. Maniatis.** 1988. 2-Aminopurine selectively inhibits the induction of β-interferon, *c-fos*, and *c-myc* gene expression. Science **240**:210–213.

Downloaded from mcb.asm.org by on March 28, 2007

Binding of the Protein Kinase PKR to RNAs with Secondary Structure Defects: Role of the Tandem A−G Mismatch and Noncontiguous Helixes†

Philip C. Bevilacqua,*,‡ Cyril X. George,§ Charles E. Samuel,§ and Thomas R. Cech‡

Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309-0215, and Department of Molecular, Cellular and Developmental Biology, University of California, Santa Barbara, California 93106-9610

Received January 15, 1998; Revised Manuscript Received March 9, 1998

ABSTRACT: The human interferon-induced double-stranded RNA (dsRNA)-activated protein kinase (PKR) is an antiviral agent that is activated by long stretches of dsRNA. PKR can also be activated or repressed by a series of cellular and viral RNAs containing non-Watson−Crick motifs. PKR has a dsRNA-binding domain (dsRBD) that contains two tandem copies of the dsRNA-binding motif (dsRBM). In vitro selection experiments were carried out to search for RNAs capable of binding to a truncated version of PKR containing the dsRBD. RNA ligands were selected by binding to His_6-tagged proteins and chromatography on nickel(II) nitrilotriacetic acid agarose. A series of RNAs was selected that bind either similar to or tighter than a model dsRNA stem loop. Examination of these RNAs by a variety of methods, including sequence comparison, free-energy minimization, structure mapping, boundary experiments, site-directed mutagenesis, and footprinting, revealed protein-binding sites composed of noncontiguous helices. In addition, selected RNAs contained tandem A−G mismatches ($^{5'AG3'}_{3'GA5'}$), yet bound to the truncated protein with affinities similar to duplexes containing only Watson−Crick base pairs. The NMR structure of the tandem A−G mismatch in an RNA helix (rGGC*AG*GCC)$_2$ reveals a global A-form helix with minor perturbations at the mismatch [Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) *Biochemistry 36,* 4449−4460]. This supports the notion that dsRBM-containing proteins can bind to RNAs with secondary structure defects as long as the RNA has an overall A-form geometry. In addition, selected RNAs are able to activate or repress wild-type PKR autophosphorylation as well as its phosphorylation of protein synthesis initiation factor eIF-2, suggesting full-length PKR can bind to and be regulated by RNAs containing a tandem A−G mismatch.

PKR is an interferon-induced human protein kinase that can regulate gene expression via multiple pathways. These include inhibition of translation initiation by phosphorylation of initiation factor eIF2α (*1, 2*) and modulation of cytokine signaling and transcription activation by the NF-κB and STAT1 factors (*3, 4*). As a result of its varied biochemical actions, PKR has been implicated as an antiviral and antiproliferative agent. For example, PKR has been shown to be a regulator of human immunodeficiency virus type 1 (HIV-1) replication (*5−7*) and adenovirus replication (*8, 9*), as well as an inducer of apoptosis (*10−13*).

PKR contains a double-stranded RNA (dsRNA)-binding domain (dsRBD) that consists of two tandem copies of the conserved dsRNA-binding motif (dsRBM) (*14−18*). The dsRBM is a 65−68 residue, compact RNA-binding motif that occurs in a large number of functionally diverse proteins from a variety of organisms (*19, 20*). The NMR structure of the dsRBM has been solved and has a secondary structure repeat α-β-β-β-α, consisting of two α-helices packed on the same face of a three-stranded antiparallel β-sheet (*20, 21*). PKR binds to dsRNA but not RNA-DNA hybrids or dsDNA. This discrimination arises because the protein makes only one energetically significant ion pair with the phosphate backbone, which is similar between the dsRNA and RNA-DNA hybrids, and instead relies on a network of 2′-OH functional groups on both strands of the helix and along the length of the binding site (*22*).

By virtue of its dsRBD, PKR can bind to and be activated by dsRNA of viral origin in a sequence-independent fashion (*21, 23−28*). The fact that viruses have evolved a variety of mechanisms for down-regulating PKR function, including production of RNAs and proteins that interfere with the activation pathway (*2, 9, 29−32*), further illustrates the importance of PKR. RNAs that regulate PKR contain numerous unpaired or non-Watson−Crick interruptions of helices, including the 3′-untranslated region (UTR) from human α-tropomyosin (*33*) and RNAs of various viruses that replicate in human cells. These include human hepatitis δ agent RNA (*34, 35*), adenovirus VA I RNA (*9, 36*), Epstein−Barr Virus EBER-1 and −2 RNAs (*37, 38*), and HIV TAR RNAs (*39−41*). PKR can also associate with ribosomes (*24*), perhaps binding to non-Watson−Crick motifs of rRNA (*42*). Observation of PKR regulation by RNAs with second-

† This work is supported by a fellowship to P.C.B. from the Jane Coffin Childs Memorial Fund for Medical Research and by a research grant to C.E.S. from the National Institute of Allergy and Infectious Diseases. T.R.C. is an investigator of the Howard Hughes Medical Institute and a Professor of the American Cancer Society.

* To whom correspondence should be addressed. Present address: Department of Chemistry, The Pennsylvania State University, University Park, PA 16802.

‡ University of Colorado−Boulder.

§ University of California−Santa Barbara.

S0006-2960(98)00113-5 CCC: $15.00 © 1998 American Chemical Society
Published on Web 04/16/1998

ary structure defects instigated this investigation into the general role of RNA structure and sequence in binding to the dsRBM.

An RNA selection approach was utilized to search for RNAs from a random library that are capable of binding to a $(His)_6$-tagged version of the dsRBD from PKR. A pool of 10^{14} RNA sequences containing a randomized region of 50 nucleotides was prepared, from which RNAs capable of binding to the dsRBD were selected. Structural analysis of these RNAs suggests a general role for non-Watson−Crick motifs, as well as noncontiguous helices, in dsRBD binding.

MATERIALS AND METHODS

PKR Protein Construct. Selection experiments were performed with H_6Tp20, an N-terminal, $(His)_6$-tagged protein (*22*) that contained residues 1−184 of the wild-type (Wt) 551-residue PKR protein (*43*). H_6Tp20 protein was overexpressed in *Escherichia coli* and purified to greater than 90% homogeneity by affinity chromatography with nickel(II) nitrilotriacetic acid agarose (*22*).

Selection Procedure. Selection procedures followed the basic approaches described (*44, 45*).

(1) Preparation of the Randomized dsDNA Template. A double-stranded DNA fragment was prepared by PCR using top- and bottom-strand primers, TS2 and BS2, respectively, to amplify a DNA template, Random2: TS2, GGGGGAATTCTAATACGACTCACTATAGGGAGAGCGGAAGCGTGCTGGGCC; BS2, GGGGGGATCCATCGACCTCTGGCTTAAG; Random2, GCGGAAGCGTGCTGGGCC-N(50)-CTTAAGCCAGAGGTCGAT; where each N position is an approximately equimolar mixture of the four nucleotides. The resultant dsDNA fragment had a 50-nucleotide random region, a T7 promoter sequence, and *Eco*RI and *Bam*HI cloning sites. The PCR reaction was carried out in a total volume of 3.3 mL using 165 pmol, or approximately 10^{14} different sequences of the Random2 template. PCR reactions were heated to 95 °C for 1 min prior to addition of 16.5 μL of 5 units/μL Taq polymerase (Perkin-Elmer). The PCR reaction was then performed for five cycles of a three-temperature amplification (1 min at 95 °C, 1 min at 54 °C, and 1 min at 72 °C), using high primer concentrations of 5 μM each. This led to a single band of correct mobility on an agarose gel. Eight and more rounds of PCR led to slower and faster mobility artifact bands and were thus avoided. These artifacts have been attributed, and least in part, to formation of heteroduplexes (*46*).

Reactions were examined on a 5% NuSieve agarose gel (FMC BioProducts), which indicated approximately 460 pmol of a double-stranded DNA product 129 bp in length. The PCR reaction was extracted with 1 vol of chloroform:isoamyl alcohol (24:1), ethanol precipitated, and washed in 70% ethanol. The resulting pellet was stored in TEN_{50} [10 mM Tris (pH 7.5), 1 mM EDTA, and 50 mM NaCl].

(2) Preparation of the Randomized RNA Library. A T7 transcription reaction was performed as described previously (*22, 47*), with the following exceptions. Transcription was in the presence of [α-^{32}P]ATP in a total volume of 1 mL, and half of the dsDNA product from the PCR reaction was used as a template. After reaction for 2 h at 37 °C, 60 μL of 0.5 M EDTA was added to chelate the Mg^{2+}. The solution was ethanol precipitated, dissolved in 100 μL of TE [10 mM Tris (pH 7.5) and 1 mM EDTA], and passed over a Sephadex G-25 spin column. The eluate was digested with 50 units of DNase I (Pharmacia) for 20 min at 37 °C. The resulting solution was phenol extracted, ethanol precipitated, washed twice in 70% ethanol, and redissolved in 100 μL of HE [10 mM Hepes (pH 7.0), 0.1 mM EDTA]. When this purified RNA was run on a denaturing 6% polyacrylamide gel with radiolabeled single-stranded DNA markers, there was one major RNA product close to the expected length of 102 nucleotides. RNA concentration was determined spectrophotometrically. This pool of RNAs is referred to as pool 0 RNA. The pool of RNAs transcribed *after* round *n* of selection is referred to as pool *n* RNA.

(3) Selection of the H_6Tp20-Binding RNAs. Selection steps were at 22 °C. Before selection for binding to H_6Tp20 protein, the RNA pool was cleared with nickel(II) nitrilotriacetic acid agarose resin to remove potential resin-binding RNAs, as follows. Internally ^{32}P-labeled RNA (100 μL) was renatured by heating at 95 °C for 3 min in TE and incubating at 22 °C for 10 min. Selection buffer (SxnB) was added [1 × SxnB: 150 mM NaCl, 25 mM Hepes (pH 7.5) and 5 mM 2-mercaptoethanol], followed by 40 μL of a 50% slurry of nickel(II) nitrilotriacetic acid agarose resin (Qiagen) previously equilibrated in 1 × SxnB. This mixture was manually shaken for 10 min at 22 °C to ensure good contact between the resin and the solution. The mixture was centrifuged for 10 s at 14 000 rpm in an Eppendorf centrifuge, and the supernatant was reserved. The resin was washed once with 100 μL of 1 × SxnB, and the wash and supernatant were combined. When the cleared resin was washed three times with 500 μL of 1 × SxnB (compared to 10 × 500 μL washes of the protein-bound resin), only 0.044% or less of the RNA remained bound to the resin. This indicates that inadvertent enrichment of resin-binding RNAs did not occur.

After the preclear steps, herring sperm DNA, as per Bevilacqua and Cech (1996), and H_6Tp20 were added at concentrations of 0.1 mg/mL and 1 μM, respectively. Binding was allowed to occur in solution for 5 min prior to introduction of 100 μL of a 50% slurry of nickel(II) nitrilotriacetic acid agarose resin in 1 × SxnB. This step allowed RNA to bind to a free protein rather than an immobilized one. Five minutes is sufficient time to permit complete binding of H_6-Tp20 to model duplex RNA (*22*). Binding was in a higher salt than in a previous study (*22*), 150 mM versus 10 mM NaCl, to increase the stringency of binding. Although the salt dependence for binding of a 20 bp model dsRNA to H_6Tp20 is shallow, the increase in salt concentration used here still corresponds to an ~10-fold decrease in binding affinity (*22*).

The RNA-protein-agarose mixture was manually shaken for 10 min at 22 °C, followed by centrifugation as above. The supernatant was removed, and the resin was washed 10 times with 500−600 μL of 1 × SxnB for each wash. The RNA was eluted with two or three 100 μL portions of 1 × elution buffer [ElB: 150 mM NaCl, 25 mM Hepes (pH 7.5), 5 mM 2-mercaptoethanol, and 250 mM imidazole]. Imidazole can directly compete with histidine for Ni^{2+}-binding sites. EDTA was not chosen as the elution agent since it has the potential for directly eluting RNAs with high Ni^{2+} affinity and leading to a false positive. The eluate fractions were pooled and phenol extracted. Glycogen carrier (20 μg from Boehringer Mannheim) was added, the RNA was

Table 1: Progress of in Vitro Evolution[a]

round	[RNA] (μM)	[protein] (μM)	%protein[b] bound
1	11	5	3.3
2	1	5	2.5
3	0.7	1	2.1
4	1	1	1.5
5	1.6	1	6.9
6	1.7	1	0.95
7	1.8	1	2.2
8	2.6	1	2.2
9	3	1	2.9
10	3.4	1	5.5
11	1.6	0.33	5.2
12	2.7	0.33	1.9

[a] The competitor was ssDNA for round 1, tRNA for all subsequent rounds. [b] Percent protein bound is equal to the percent RNA eluted times the fold-excess of RNA concentration over protein concentration. Liquid scintillation counting was used to determine the amounts of eluted and total RNA. Percent RNA eluted is equal to the ratio of the amount of the eluted RNA to the amount of the total RNA, $\times 100$.

ethanol precipitated and washed with 70% ethanol. The pellet was dried in a speed-vac, dissolved in 50 μL of TE, and passed over a Sephadex G-25 spin column. The RNA was ethanol precipitated, washed with 70% ethanol, dried in a speed-vac, and dissolved in 10 μL of TE. Fractions were collected at all steps and quantitated via liquid scintillation counting.

(4) RT-PCR To Convert the Selected RNA into dsDNA Template. Selected RNA (4 μL) was annealed with BS2, and cDNAs were obtained by reverse transcription for 30 min at 60 °C using AMV reverse transcriptase (Life Sciences). To this solution was added 50 μL of a master mix containing buffer and TS2. This mixture was heated at 95 °C for 3 min prior to addition of Taq polymerase. PCR was performed for 10 cycles, each involving the two-temperature amplification: 1 min at 95 °C and 1 min at 72 °C. The products of the RT-PCR reaction were examined on a 5% agarose gel. Two control reactions were carried out in parallel with each RT-PCR reaction. One control was without addition of selected RNA to test for DNA contamination of the stock solutions. The other control was with 1 μL of RNA added but without any reverse transcriptase to test for DNA contamination of the selected RNA.

(5) Later Rounds of Selection. Subsequent rounds of selection were the same as described for round 1 with the following exceptions. T7 transcription reactions were in 100 μL volumes and loaded directly on the Sephadex G-25 spin columns without any precipitation. The DNase step was not performed at this time, but rather after selection (see below). The RNA was refolded at 95 °C for only 1 min, and the competitor was 0.1 mg/mL ($\sim$6 μM) $tRNA^{Phe}$, not 0.1 mg/mL single-stranded (ss) DNA, to increase the stringency of binding (*22, 48*). tRNA can compete 2−8-fold better than ssDNA for H_6Tp20 binding to TAR and dsTAR, respectively (*22*). The concentration of H_6Tp20 was lowered in later rounds to increase the stringency of binding (see Table 1). After the selected RNA was ethanol precipitated with glycogen carrier, it was subjected to a DNase digestion. This digestion was in 40 μL at 37 °C for 2 h with 40 units of RQ1 RNase-free DNase (Promega). The DNase step was moved so that it followed the selection, thereby reducing the amount of nucleic acid present during the digestion. After the DNase step, the solution was phenol extracted. Glycogen (20 μg) was added, and the RNA was ethanol precipitated and washed with 70% ethanol. The pellet was dried in a speed-vac and dissolved in 10 μL of TE. PCR steps involved 15−23 rounds of the above two-temperature amplification.

After 12 rounds of selection, the RNA was converted into cDNA by reverse transcription and amplified by PCR. The PCR fragments were digested with *Eco*RI and *Bam*HI and cloned into the plasmid pUC19. Sequences of isolated clones were determined by dideoxy sequencing, using sequencing gels that contained 40% formamide and 7 M urea in order to resolve compressions.

Preparation of RNAs and DNAs. RNA was prepared from desired clones by runoff T7 transcriptions. The appropriate plasmid was digested with *Bam*HI and transcribed as described previously (*22*). The reaction was ethanol precipitated, run on a 6% denaturing polyacrylamide gel, and product bands visualized by UV shadowing. RNA was recovered by a crush and soak method (*22*).

All other RNA and DNA were prepared by solid-phase synthesis and deblocked as previously reported (*49, 50*). RNA sequences are found in the appropriate figure or figure caption.

Binding Assay. Binding affinity between H_6Tp20 and 5′-^{32}P-labeled RNA was determined by native-gel mobility-shift assays, as previously described (*22*). Fits are to eq 1,

$$\theta = \epsilon \frac{[PKR]}{[PKR] + K_d} \quad (1)$$

where θ is the fraction of RNA bound, ϵ is the observed maximum fraction bound, and K_d is the dissociation constant (*22*). As necessary, complexes 1 and 2 were summed to give total RNA bound and obtain a K_d for binding of one protein to free RNA. This appears justified because complex 2 does not begin to form until essentially all of the free RNA is shifted. This suggests that binding of the first and second proteins are not strongly coupled, and that conversion of complex 1 into complex 2 does not effect the observed K_d for complex 1. Some of the gel mobility-shift data was cooperative and not well-described by a simple hyperbola. The origin of this effect is not well understood. In these cases, the K_d reflects the concentration of protein required to give half-maximal RNA bound.

Structure Mapping. RNAs prepared by runoff T7 transcription were treated with calf intestinal phosphatase (CIP) to remove the 5′-triphosphate, reacted with polynucleotide kinase and [γ-^{32}P]ATP, repurified by gel electrophoresis, excised from the gel, eluted overnight in TEN_{250} at 4 °C, ethanol precipitated, and resuspended in TE. 5′-Labeled RNA in TE was renatured by heating at 95 °C for 1 min and incubating at 22 °C for 10 min. RNA was digested with the appropriate nuclease for 1 h at 22 °C, under native conditions in 1 × SxnB. Nuclease concentrations were chosen to give limited hydrolysis, and 1 x concentrations were 1×10^{-5} units/μL RNase T1, 2×10^{-3} units/μL RNase U2, 2×10^{-6} units/μL RNase A, 1×10^{-4} units/μL RNase T2, and 0.07 units/μL RNase V1. Samples were either loaded immediately on a denaturing polyacrylamide gel or frozen on dry ice, thawed on ice, and loaded on a gel. Samples treated with imidazole were as previously described (*51*). Briefly, a 10 μL aliquot of 5′-end labeled RNA was mixed with an equal volume of 4 M imidazole (pH 7.0) and

incubated at 22 °C for 16 h. A 10 μL portion of this mixture was heated at 90 °C for 10 min as a denaturing control. To each aliquot were added 10 μL of 0.3 M NaOAc (pH 5.0) and 200 μL of a 2% $NaClO_4$ solution in acetone. The RNA was pelleted by centrifugation, washed with 400 μL of acetone, air-dried, and dissolved in 10 μL of TE followed by 10 μL of a 2× formamide loading buffer.

Sequencing lanes for A and G were prepared by limited hydrolysis with RNases U2 and T1, respectively, under denaturing conditions. Control lane was same as the G-sequencing lane with RNase T1 omitted. All-nucleotide lanes were prepared by treatment with alkali (*52*).

Footprinting Experiments. RNAs were 5′-end labeled, purified, and renatured as in the structure-mapping experiments. An aliquot of renatured RNA was incubated for 5 min at 22 °C with 7.7 μM H_6Tp20, and then incubated with the appropriate nuclease for 30 min at 22 °C. Samples were placed on ice, mixed with an equal volume of a formamide/0.5% SDS loading buffer, and 1 μL was loaded on an 8% polyacrylamide gel.

Boundary Experiments. RNA was either 5′-end labeled as in the structure mapping experiments or 3′-end labeled with poly(A) polymerase and [α-^{32}P]cordycepin-5′-triphosphate as per manufacturer's instructions (United States Biochemical). Labeled RNAs were treated with alkali to generate a ladder of sequences. Hydrolyzed RNAs were mixed with 5 μM H_6Tp20 and loaded on a native gel, as previously described (*22*). Bound and unbound RNAs were visualized by autoradiography, excised from the gel, and eluted by a crush and soak method. RNAs were ethanol precipitated and run on an 8% denaturing polyacrylamide gel, along with T1 and alkali digests.

In Vitro PKR Activation Experiments. The effect of H_6-Tp20-selected RNAs on PKR kinase activity was determined by measurement of both PKR autophosphorylation and eIF-2 phosphorylation, as previously described (*53, 54*). PKR-(Wt) enzyme was purified from ribosomal salt-wash fractions prepared from interferon (IFN)-treated human amnion U cells as described (*53*), except that the procedure was modified to include a MonoQ ion-exchange fractionation step (*54*). [γ-^{32}P]ATP-mediated PKR autophosphorylation catalyzed by PKR(Wt) was carried out in the absence of added activator or in the presence of RNA as indicated. For measurement of eIF-2α phosphorylation by PKR, 0.25 μg of purified eIF-2 was added to the standard reaction mixture. The ^{32}P-labeled products were analyzed by SDS−polyacrylamide gel electrophoresis and autoradiography. The procedures have been described in detail (*53*). Quantitation was carried out using a Bio-Rad Molecular Imager model GS525.

RESULTS

Use of Ni^{2+}-Affinity Chromatography to Select Protein-Binding RNAs. Several control experiments were performed to determine optimal conditions for selection experiments. Binding of complexes of H_6Tp20 with [5′-^{32}P]TAR or with dsTAR complex (*22*) to nickel(II) nitrilotriacetic acid agarose resin in batch (Materials and Methods) was compared to binding in a silica-based spin-column (Qiagen). [dsTAR is a double-stranded version of TAR with a 24 bp stem in which the three TAR bulges are deleted and G:U wobble pairs are converted to G:C base pairs (*22*).] In general, 10% more of the ^{32}P-labeled RNA was bound in batch compared to spin-column mode, as determined by liquid scintillation counting (data not shown). More importantly, in batch mode trials using only two 600 μL washes, only 0.1% of the RNA was bound and then eluted from the column without protein present, compared to 17−33% of the RNA with protein present. For the spin columns, 3−7% of the RNA was bound and then eluted from the column without protein present, compared to 10−23% of the RNA with protein present. Such a high background of RNA binding to the spin column in the absence of protein was unacceptable.

During each of the batch-mode selection steps, the protein dependence of binding was determined by comparing the counts per minute of the 300 μL of 250 mM imidazole-containing eluate to the counts per minute of the 10th and final 500 μL wash of the RNA-protein-agarose mixture. Between 91 and 98% more counts per minute eluted upon addition of the imidazole-containing buffer, indicating that the elution was specific to protein-bound RNAs. Also during each of the batch-mode selection steps, the efficiency of the elution was determined by comparing the counts per minute of the 300 μL of eluate to the counts per minute of a suspension of the used resin. This percent efficiency varied from 92 to 99% depending on the round of selection. In conclusion, nickel(II) nitrilotriacetic acid agarose resin appears to provide a generally useful approach for selection of RNAs that bind to a $(His)_6$-tagged protein. When used in batch mode, the resin does not significantly bind nucleic acids, while specific and efficient elution of protein and its bound RNAs is achievable.

Selection for RNAs That Bind the dsRBD from PKR. A library of $\sim 10^{14}$ different sequences was generated by overlap extension of top- and bottom-strand primers with a DNA template that had a stretch of 50 randomized positions, followed by transcription with T7 RNA polymerase. The renatured RNA pool was cleared with nickel(II) nitrilotriacetic acid agarose resin and recovered. The RNA was bound to protein in the presence of a competitor and washed exhaustively. RNA was eluted from the resin by several incubations with an imidazole-containing buffer.

Multiple rounds of selection were performed. The stringency of the binding step was increased at several points during the selection (Table 1). After 12 rounds of selection, the RNA was 5′-end labeled and its affinity for H_6Tp20 was quantitated by a native-gel mobility-shift assay (Figure 1). This heterogeneous collection of RNAs bound H_6Tp20 with an apparent K_d of 0.1 μM, 8- and 64-fold more tightly than dsTAR and TAR, respectively (Figure 1). Controls showed that pool 0 RNAs did not bind H_6Tp20 over the same range of protein concentration. At this point, selection was stopped and sequences were determined.

The RNA was converted to cDNA by reverse transcription, amplified by PCR, cloned, and sequenced. Of the 44 sequences determined, 11 were different, and five of the 11 had a 17-nucleotide consensus, "cugcaucgacaryu-ggy" where r is one of the purine bases A or G, and y is one of the pyrimidine bases C or U. These five sequences were assigned to family 1 (Figure 2). Many of the selected RNAs had deletions in the randomized region that presumably arose during the RT-PCR steps (*55*). We chose clones 2 and 11, the most frequent members of family 1, to characterize further. RNA derived from other clones, including clone 3,

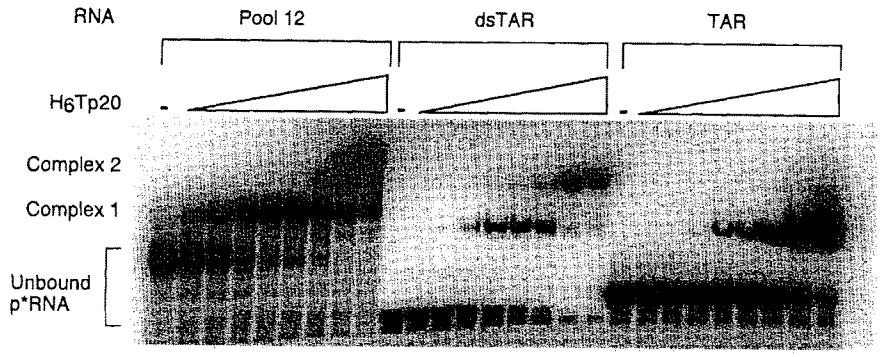


FIGURE 1: Native gel mobility shift for H_6Tp20 binding to RNA. Native-gel mobility-shift experiment for H_6Tp20 binding to trace amounts of 5′-^{32}P-labeled RNA. Pool 12 is the mixture of RNAs after 12 rounds of selection; TAR and dsTAR are partially mismatched and fully double-stranded RNA substrates, respectively, for H_6Tp20 binding as previously reported (*22*). Experiments were in the presence of 0.1 mg/mL $tRNA^{Phe}$. Native 10% polyacrylamide gel is shown. Protein binding to dsTAR resulted in slight formation of a second complex termed complex 2. Concentrations of H_6Tp20 used were 0, 0.05, 0.1, 0.25, 0.5, 1, 2, 5, and 10 μM.

	Clone		Number	Kd (μM)
		CONSENSUS cugcaucgacARyu-ggy		
Family 1	c2:	caugagguc-uau---cacugcaucgacAGcuaggcggacaaug	7	2.3
	c20:	caugaggucguaugcgaucugcaucgacAGcuaggcggagcuaucg	1	4.1
	c11:	ucggaggugg-uucagcucugcaucgacAGuu-ggc	11	5.3
	c14:	guaaggg-uuca-cucugcaucgacAAcu-gca	3	6.2
	c4:	uguuggcuugg-cucugcaucgacAGgu-ggcuu	2	6.6
	c3:	aggauggaucgagcgcuggagucuggagaaugucac	6	0.074
	c6:	ugcgauaaggugguagcgacagugauagaagc	4	2.8
	c5:	ugcacggaaaugcagaaggcgugguccgugggaggcgacaagcaucaaug	4	0.11
	c21:	cac**acgguggg-ugu**guauuacagcugacuguuggcaagcacugugaggac	3	N.D
	c32:	uau**acgguggggaugug**cccaguggggacacaugcaccguguuaguuuugca	2	N.D
	c26:	cuuaacgugaaggcgguaggggugacggacugagcgagcgccaugcauug	1	N.D

FIGURE 2: Sequence alignment of selected RNAs. Shown are sequences from the variable region only. Gaps in the alignment are represented by a dash. Numbers are assigned to clones based upon the order in which the first occurrence was sequenced. The number of identical clones found in the 44 sequences is shown on the right. The consensus stretch of 17 nucleotides is shown at the top of the figure and is boxed. The first five clones have an obvious match to the consensus and are assigned to family 1. The other clones do not have an obvious match. Another region of similarity in these clones is shown in boldface and lowercase. Y = C or U, R = A or G, and the selected region's contribution to the tandem mismatch is shown in bold capital letters. K_ds are in units of micromolar and were determined by a native-gel mobility-shift method.

bound H_6Tp20 more tightly than those of family 1 (Figure 2). Clone 3 RNA, however, did not readily afford an H_6Tp20 footprint and was not easily classified by comparative sequence analysis; consequently, its structure was not pursued in this study.

RNA derived from clones 2 and 11 bound to H_6Tp20 both on the column and by gel shifts. Dissociation constants for clones 2 and 11 RNA binding to H_6Tp20 by native-gel mobility-shift experiments are 2.3 and 5.3 μM, respectively (Figure 2). These values are similar to that for dsTAR of 0.9 μM (Figure 1).

Effect of Selected RNAs on PKR Kinase Activity. The PKR kinase purified from human cells is highly dependent upon dsRNA for activity as measured both by autophosphorylation of PKR and by the phosphorylation of protein synthesis initiation factor eIF-2 by activated PKR (*53, 54*). The ability of RNA from clones 2, 3, 11, and 14 to activate PKR was examined (Figure 3). Among these RNAs, clone 2 RNA activated PKR most efficiently. Although clone 3 RNA bound PKR more tightly than clone 2 RNA, clone 3 RNA was a much poorer kinase activator than clone 2 RNA. Optimal PKR activation with clone 2 was obtained between 0.1 μM and 1.0 μM RNA; this activation by clone 2 RNA approached the maximal activation of PKR obtained with synthetic poly(rI)-poly(rC) dsRNA (Figure 3A). Family 1 RNAs differed in their ability to activate PKR in the order clone 2 > clone 14 > clone 11. Relative to the activation of PKR obtained in the presence of clone 2 RNA, 54% activation was obtained with clone 14 RNA, 17% with clone 11 RNA, and 14% with clone 3 RNA (all RNAs tested at 1.0 μM).

To test whether selected RNAs could antagonize the activation of PKR kinase by dsRNA, purified PKR was incubated in the presence of selected RNA and then the reaction mixture was supplemented with 0.3 μg/mL (=2.4 nM) poly(rI)-poly(rC). As shown by the Figure 3B autoradiogram, clone 3 RNA was an effective antagonist of PKR activation by poly(rI)-poly(rC). About 90% reduction in PKR activation was obtained with 10 μM. Neither clone 2 nor clone 14 RNA significantly antagonized PKR activation and autophosphorylation relative to clone 3 RNA, whereas clone 11 displayed an intermediate antagonist activity.

Sequence Alignment and Free-Energy Minimization of Selected RNAs. The 17-nucleotide consensus sequence common to family 1 clones (Figure 2) can form 11−12 Watson−Crick or G−U wobble pairs with the fixed sequence 3′-primer-binding site. The proposed base-paired region is interrupted by an internal tandem A−G mismatch (with clone 14 forming an internal-5′AA3′/5′AG3′ mismatch). Each A−G pair in a tandem A−G mismatch can form an imino two hydrogen-bonded conformation, with A-NH6 to G−O6 and A−N1 to G-NH1 (imino) (*56*), which when added to

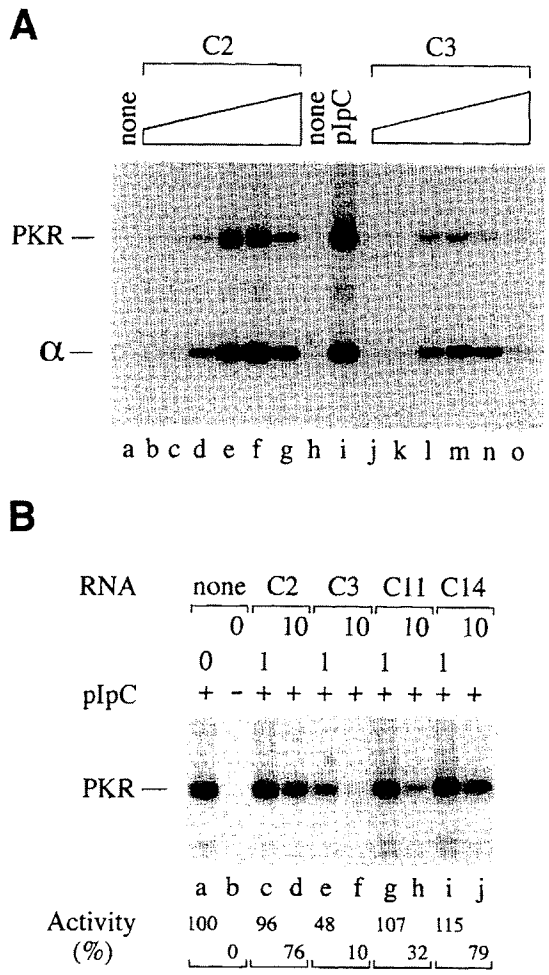


FIGURE 3: Effect of Selected RNAs on PKR Kinase Activity. (A) In vitro activation of PKR by selected RNAs. Shown is an autoradiogram of a 10% SDS−polyacrylamide gel. RNA samples, either clone 2 (lanes b−g) or clone 3 (lanes j−o), were examined for their ability to activate purified PKR kinase. RNA was added as indicated to the standard reaction mixture which contained [γ-^{32}P]-ATP and eIF-2. Concentrations of clone 2 and 11 RNAs used were: lanes b and j, 0.0001 μM; lanes c and k, 0.001 μM; lanes d and l, 0.01 μM; lanes e and m, 0.1 μM; lanes f and n, 1.0 μM; lanes g and o, 10 μM RNA. Reactions carried out in the absence of RNA (lanes a and h) or in the presence of the optimal concentration of poly(rI)-poly(rC) (lane i) are shown as controls. PKR denotes the position of ^{32}P-labeled PKR, and α denotes the position of ^{32}P-labeled α-subunit of eIF-2. (B) Antagonism of PKR activation by selected RNAs. Shown is an autoradiogram of a 10% SDS−polyacrylamide gel. PKR was incubated with RNA samples of clone 2 (lanes c and d), clone 3 (lanes e and f), clone 11 (lanes g and h), and clone 14 (lanes i and j) at either 1.0 or 10 μM prior to the addition of 0.3 μg/mL (=2.4 nM) poly(rI)-poly(rC). Reactions carried out in the absence of added RNA (lane b) or in the presence of the optimal concentration of poly(rI)-poly(rC) (lane a) but without added selected RNA, are shown as controls. PKR denotes the position of ^{32}P-labeled PKR. The autophosphorylation activity of PKR kinase was quantified and is expressed as a percent of the activity observed in the presence of an optimal concentration of poly(rI)-poly(rC).

the flanking helical regions would give a dsRNA-like helix of length 13 bp for RNA clones 4 and 14 and 14 bp for RNA clones 2, 11, and 20. This is close to the minimal length for binding of 16 bp identified in a previous study (*22*). These data suggest the model that the 3′-stem-loop of the family 1 RNAs forms the structure shown in Figure 4, containing the tandem A−G mismatch.

Secondary structures of various RNA clones were predicted by free-energy minimization using *Mfold* v2.3 (*57, 58*). Clone 11 RNA was predicted to form the structure in Figure 4 containing two major stem loops, P2/L2 and P3/L3. We show below that structure mapping, boundary experiments, site-directed mutagenesis, and footprinting experiments are consistent with this predicted secondary structure.

Clone 2 RNA was also predicted to form a structure with two major stem loops, P2/L2 and P3/L3. However, the 3′-stem-loop predicted to have the lowest free energy structure (−29.4 kcal/mol) has a structure different from that shown in Figure 4. The computer-predicted structure has an asymmetric loop with a top strand A opposing a bottom strand 5′CAGAG3′. Use of a window size of 0 in *Mfold* (*57*) identifies a suboptimal secondary structure only 0.7 kcal/mol less stable (−28.7 kcal/mol) with the tandem A−G mismatch structure shown in Figure 4. The lowest free-energy structure positions a nonconserved A in a stem base pair, while the suboptimal structure positions this A in a bulge as shown in Figure 4. It is most consistent to have a nonconserved residue not making an important interaction, further supporting the tandem A−G-containing structure of Figure 4.

Structure Mapping of Selected RNAs. To test whether the selected RNAs fold into the secondary structures suggested by sequence alignment and free-energy minimization, the RNAs were subjected to structure mapping. Cleavage of RNA clones 2 and 11 under native conditions by RNases U2, T1, A, T2, and V1, and by imidazole, is shown in Figure 5 and summarized in Figure 4. RNases U2, T1, A, and T2 are single-strand-specific nucleases that leave 3′-phosphate or 2′,3′-cyclic phosphate termini (*59*). RNase V1 is a double-strand specific nuclease that cleaves RNA without base specificity to leave a 5′-phosphate terminus (*59*). RNase V1 can also cleave single-stranded regions that exist in a stacked arrangement (*59, 60*). High concentrations of imidazole (pH 7.0) preferentially cleave RNA in single-stranded regions, leaving a 3′-phosphate terminus (*51, 61*). Positions that were cleaved by both single-stranded-specific RNases and by RNase V1 are not assigned a specific cleavage arrow in Figure 4. These ambiguous cleavages may be caused by single-stranded residues that are stacked, or by two competing RNA conformations that are in equilibrium.

The RNase and imidazole cleavage patterns were largely consistent with the predicted secondary structures (Figure 4). RNase V1 cleaved the selected RNAs within predicted double-stranded regions or at nucleotides one to two positions 3′ of double-stranded regions. A 3′-dangling end is able to stack on its proximal helix (*62*), consistent with its cleavage by RNase V1. Interestingly, RNase V1 cleaved clone 11 RNA lightly after position A52 and G53 of the tandem A−G mismatch. Single-strand-specific RNases and imidazole cleaved the selected RNAs within predicted single-stranded regions or at nucleotides involved in base pairs at the end of a stem. Such base pairs may be partially frayed, leaving them susceptible to attack.

As mentioned, the structure for clone 2 RNA shown in Figure 4 is a suboptimal structure according to calculation. The lowest free-energy structure predicted for clone 2 RNA

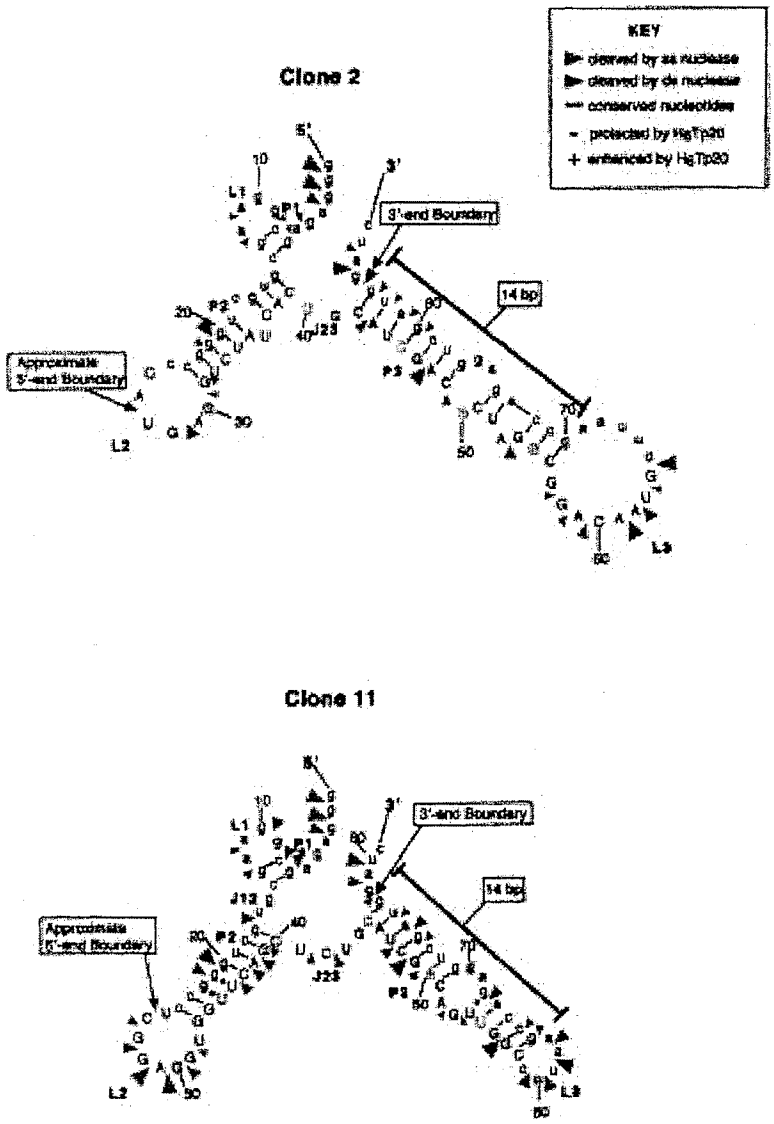


FIGURE 4: Secondary structural models for selected RNAs. Positions of cleavage by single-stranded probes (RNase U2, RNase T1, RNase A, RNase T2, and imidazole) are shown using red arrowheads with the size of the arrowhead proportional to the intensity of the cleavage band. Positions of cleavage by the double-stranded probe RNase V1 are shown using green arrowheads with the size of the arrowhead proportional to the intensity of the cleavage band. Ambiguous sites are not marked with any arrowhead. Nucleotides are numbered every 10 residues with each fifth residue outlined. Positions of the 5′- and 3′-boundaries are shown, as are positions of H_6Tp20 protection and enhancement of nuclease cleavage. Conserved nucleotides are bold and colored purple, and the length of the 3′-stem is indicated. Length of the tandem-mismatch containing stem is labeled. Number of base pairs includes Watson−Crick base pairs denoted with a dash, G−U wobble pairs denoted with a dot, and the tandem mismatch. Fixed-sequence primer-binding sites are in lower case, and nucleotides from the random region are in upper case. Positions of nuclease cleavage protection and enhancement by H_6Tp20 are indicated with minus (−) and plus (+) signs, respectively. Paired, joining, and loop regions are denoted with a P, J, and L, respectively, and numbered in order of occurrence from the 5′- to 3′-direction.

had two inconsistencies with the cleavage pattern of Figure 4: positions 63 and 64, predicted to be double stranded (data not shown), were in fact cleaved by RNases T2 and T1, respectively. Thus, the nuclease-mapping data are consistent with the tandem A−G containing secondary structures shown in Figure 4, rather than the calculated optimal structure.

Boundary Experiments on Selected RNAs. Clone 2 and clone 11 RNAs were tested to define the minimal set of nucleotides required for protein binding. RNAs were either 5′- or 3′-end labeled, hydrolyzed to give a ladder of sequences, and mixed with protein. Bound and unbound RNAs were separated by native gel electrophoresis, concentrated, and run on a sequencing gel. 5′-End-labeled sequences show a well-defined cutoff between bound and unbound sequences with only four residues dispensable at the 3′-end (Figures 4 and 6). Identical experiments were carried out on clone 2 RNA (data not shown). This cutoff retains the 3′-stem-loop and confirms its importance in H_6-Tp20 binding.

3′-End-labeled sequences show a poorly defined cutoff between bound and unbound sequences, with bound sequences gradually decreasing in intensity with a decrease in

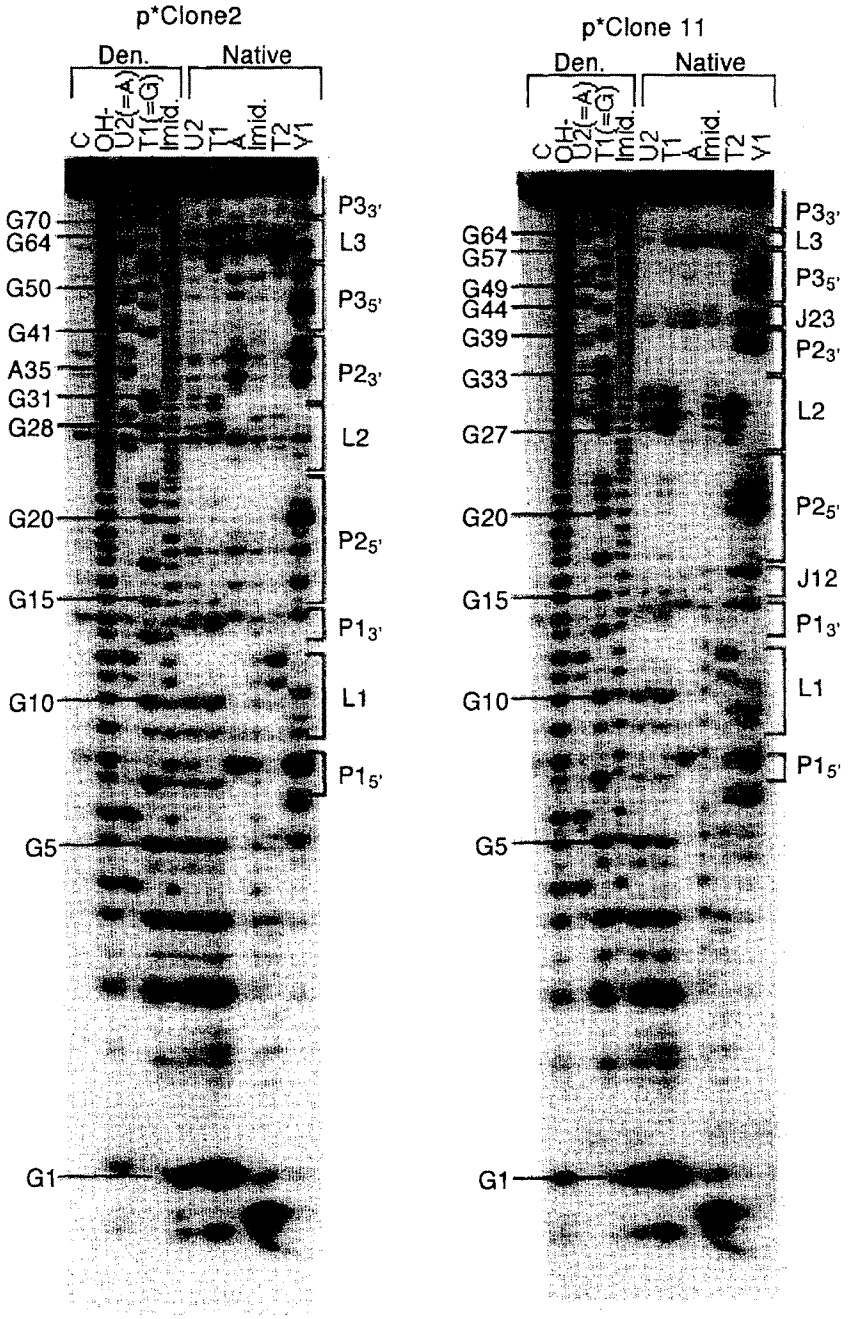


FIGURE 5: Structure mapping of selected RNAs. Denaturing 16% polyacrylamide gel is shown. RNA samples derived from clones 2 and 11 were 5′-^{32}P-labeled. Left-hand five lanes are under denaturing conditions (denoted "Den."), and right-hand six lanes are under native conditions (denoted "Native"). Labels for lanes are as follows: C is a control (no nuclease) sample, OH− is a limited alkaline digest; U2, T1, A, T2, and V1 are limited digests under native conditions with RNases U2, T1, A, T2, and V1, respectively; Imid. is a limited imidazole digest under native conditions. Samples were also run on a denaturing 8% polyacrylamide gel to help separate and identify the 3′-most cleavages (data not shown). Paired, joining, and loop regions are denoted with a P, J, or L, respectively, and the 5′- or 3′-strand of a region is denoted with a subscript.

RNA length and unbound sequences gradually increasing in intensity with a decrease in RNA length (Figure 6). As such, it is not possible to define precisely a 5′-end for RNA clone 2 or 11 that is absolutely required for binding. It is clear from the 3′-end labeled RNA boundary experiment, however, that the 14 bp 3′-stem-loop with the tandem A−G mismatch is not capable by itself of binding to H_6Tp20. The shortest sequences that are approximately 50% bound begin with residues 27 and 26 for RNA clones 2 and 11, respectively (Figure 6). RNAs derived from clones 2 and 11 that begin with residues 27 and 26, respectively, are capable of forming minihelices at their 5′-ends of 3 and 5 bp in length,

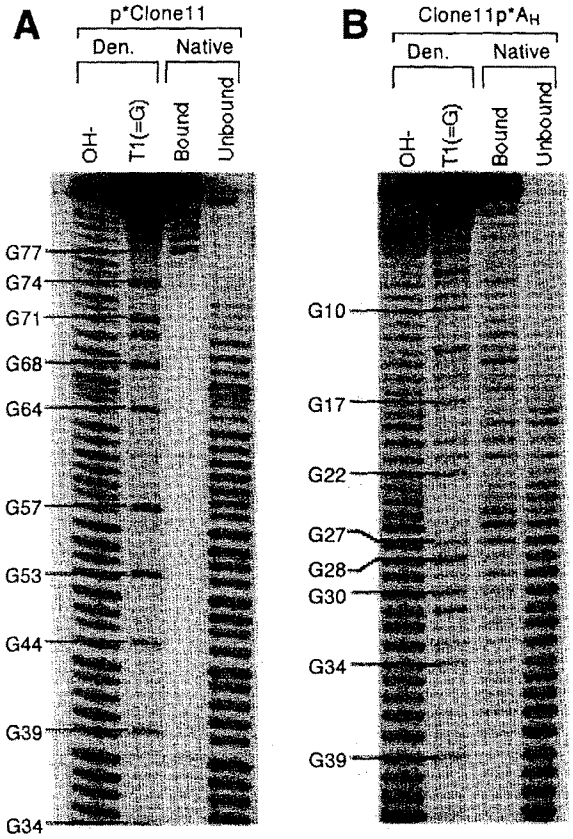


FIGURE 6: Boundary experiment for clone 11 RNA. Denaturing 8% polyacrylamide gel is shown. RNA was either (A) 5′-^{32}P-labeled, denoted "p*Clone11", or (B) 3′-^{32}P-cordycepin-labeled, denoted "Clone11p*A_H". Left-hand two lanes contain samples treated under denaturing conditions (denoted "Den."), and right-hand two lanes contain samples treated under native conditions (denoted "Native"). Labels for lanes are as follows: OH− is a limited alkaline digest; T1 is a limited digest with RNase T1; "Bound" and "Unbound" are RNA fragments from the limited alkaline digest that either bound or did not bind, respectively, to 5 μM H_6Tp20. Samples were also run a shorter distance on the 8% polyacrylamide gel to examine shorter fragments (data not shown).

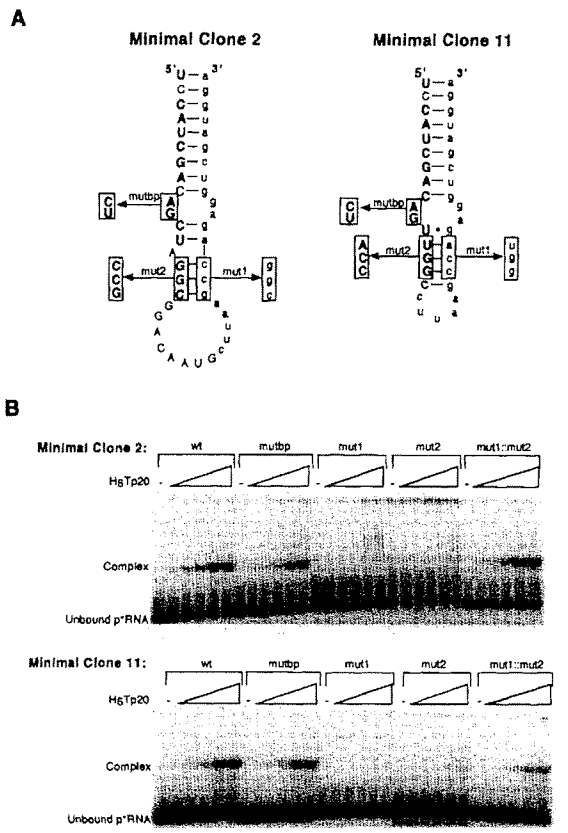


FIGURE 7: Structure and function of minimal-length RNAs. (A) Secondary structures for minimal RNAs derived from clones 2 and 11 consistent with structure mapping experiments, site-directed mutagenesis experiments, and free-energy minimization using *Mfold* v2.3 (*57*). Conserved nucleotides are bold. Positions of site-directed mutations in mutbp, mut1 and mut2 are indicated; mut1::mut2 is the double mutant with changes simultaneous for mutant 1 and mutant 2. (B) Native-gel mobility-shift experiments for H_6Tp20 binding to trace amounts of 5′-^{32}P-labeled RNA. Protein binding to RNA in the presence of 0.1 mg/mL $tRNA^{Phe}$ resulted in formation of a single complex. Concentrations of H_6Tp20 used were 0, 0.5, 0.1, 0.2, 5, and 10 μM.

respectively, as evaluated by free energy minimization using *Mfold* v2.3 (*57*). These minihelices are predicted to fold in the presence of the 14 bp 3′-stem-loop structures shown in Figure 4. Since the 14 bp 3′-stem-loop itself cannot bind protein, and the minimal-length stem that can bind H_6Tp20 is known to be 16 bp (*22*), it is plausible that the 3−5 bp 5′-minihelices augment the A−G containing 3′-stem-loop stretch to form helical segments of sufficient length for protein binding. This is consistent with H_6Tp20 footprints present in both the 3′-stem-loop and the 5′-region of full-length clone 11 RNA (Figures 4 and 8, see below).

Minimal-Length RNAs: Structure and Site-Directed Mutagenesis. To test the secondary structure of the 3′-stem-loop and the importance of the tandem A−G mismatch in H_6Tp20 binding, a series of minimal RNAs and site-directed mutants was prepared. The minimal RNAs consist of the 3′-stem-loop with two additional base pairs at the end of the stem to give a 16 bp substrate (Figure 7A). Four mutant RNAs were also constructed: the base-pair mutant converts the tandem A−G mismatch of the model to two Watson−Crick base pairs, mutants 1 and 2 disrupt the stem proximal to the loop, and the double mutant combines mutants 1 and 2 to restore the proximal stem in the model (Figure 7A).

The wild-type, base-pair, and double-mutant RNAs were all capable of binding to H_6Tp20, while mutants 1 and 2 (in which the stem proximal to the loop was disrupted) were not capable of binding H_6Tp20 (Figure 7B). These results provide evidence that the secondary structures shown in Figure 7A are correct, that the tandem A−G mismatch is roughly equivalent to a Watson−Crick base pair in its ability to bind protein, and that the stem proximal to the loop is necessary for binding H_6Tp20. In addition, the sequence of the clone 2 double mutant, which is capable of forming the tandem A−G mismatch structure (Figure 7A), is not capable of forming the alternative secondary structure predicted by *Mfold* (*57*) to have an asymmetric loop. Since the double mutant retains H_6Tp20 binding capability, the secondary structure for clone 2 RNA shown in Figure 7A is capable of binding the protein.

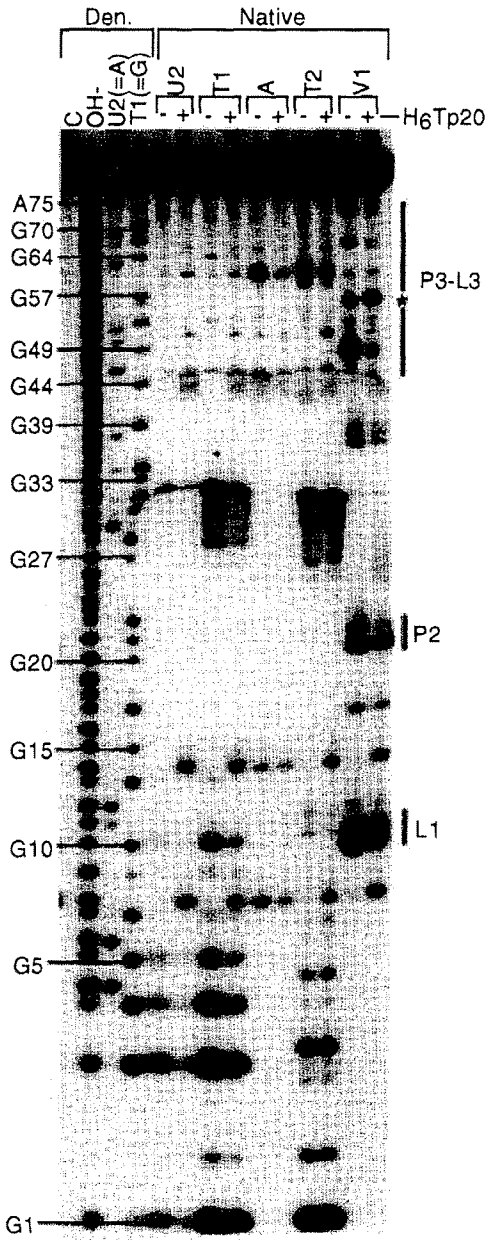


FIGURE 8: Footprinting experiment for clone 11 RNA. Denaturing 8% polyacrylamide gel is shown. RNA was 5′-^{32}P-labeled. Lanes are labeled as in Figure 5. In indicated lanes, H_6Tp20 was added at 7.7 μM (enough to give complete mobility shift of the complex prior to nuclease digestion). Nucleotides are labeled along the left-hand side of the gel. To the right of the gel, regions of H_6Tp20-dependent protection from nuclease cleavage are denoted with a bold line, and a site of H_6Tp20-dependent enhancement of cleavage is denoted with an asterisk (*).

Interaction of Selected RNAs with the dsRBD from PKR. To examine the binding sites for H_6Tp20 on the selected RNAs, footprinting experiments were performed. Treatment of the clone 2 and 11 RNAs with RNases U2, T1, A, T2, and V1 in the presence and absence of saturating H_6Tp20 is shown in Figure 8 and summarized in Figure 4. Positions that were cleaved by both single-stranded-specific RNases and by RNase V1 are not assigned a protection or enhancement. H_6Tp20 binding to clone 11 RNA gave a series of well-defined protections from RNase V1 in double-stranded sections, as well as one enhancement of RNase V1 cleavage at G56. In addition, protection from single-stranded nucleases occurred in L3.

H_6Tp20 binding to clone 2 RNA yielded only one well-defined protection, occurring with RNase V1 at position 80 in the 3′-stem. In sum, footprints map largely to the 3′-stem-loop of the structures shown in Figure 4, emphasizing the importance of this stem and its secondary structure.

Minimal-Length RNAs: Effect on PKR Kinase Activity. The relationship between RNA-binding activity of the minimal-length RNAs to H_6Tp20 and their ability to activate PKR kinase activity was examined. The clone 2 and 11 series of minimal RNAs and their site-directed mutants were tested at varying concentrations for their ability to activate PKR as measured by the phosphorylation of eIF-2α. As shown in Figure 9, the minimal-length clone 2 RNA activated PKR, but much less efficiently than the parental clone 2 RNA and only at relatively high concentrations. Minimal-length clone 2 mutant RNAs, like the minimal-length clone 2 RNA, were also very poor activators of PKR kinase activity (data not shown).

The ability of the minimal-length RNAs to antagonize the activation of PKR kinase by dsRNA was also examined. Purified PKR was incubated in the presence of minimal RNA, either wild-type or mutant, and then the reaction mixture was supplemented with poly(rI)-poly(rC). A series of clone 2 and 11 minimal RNAs and site-directed mutants was examined (Figure 10). As shown by Figure 10, minimal-length clone 2 base-pair mutant was a very effective antagonist of PKR activation. Mutant 1 minimal-length clone 2 RNA was a poor antagonist, whereas wt, mutant 2 and the double-mutant minimal-length clone 2 RNAs showed an intermediate ability to antagonize activation. By contrast, the minimal-length clone 11 series of RNAs, either wild-type or mutant, did not significantly antagonize PKR activation (data not shown).

DISCUSSION

The dsRBM is a conserved set of amino acids that constitutes a compact folding domain capable of binding dsRNA (*19*−*21*, *23*). In addition, dsRBM-containing proteins bind to RNAs containing single-stranded or non-Watson−Crick secondary structures (*9*, *28*, *33*−*41*, *63*). We have performed in vitro selection experiments to search for RNA sequences and structures that bind to the dsRBD from PKR. The selected RNAs might have been an uninteresting collection of double-stranded sequences. Instead the selected RNAs contain several non-Watson−Crick features, including a tandem A−G mismatch in the context of a dsRNA background, and a protein-binding site composed of non-contiguous helices. Thus, the results reveal new information about the binding specificity of the dsRBD from PKR.

Selected RNAs Bind H_6Tp20 and Activate PKR. After 12 rounds of selection, the pool 12 RNAs bound to H_6Tp20 approximately 8-fold more tightly than dsTAR, a model stem-loop containing 24 contiguous Watson−Crick base pairs (Figure 1). In addition, H_6Tp20 binding to selected pool 12 RNA clones was revealed by H_6Tp20-dependent protections from RNase cleavage (Figure 8).

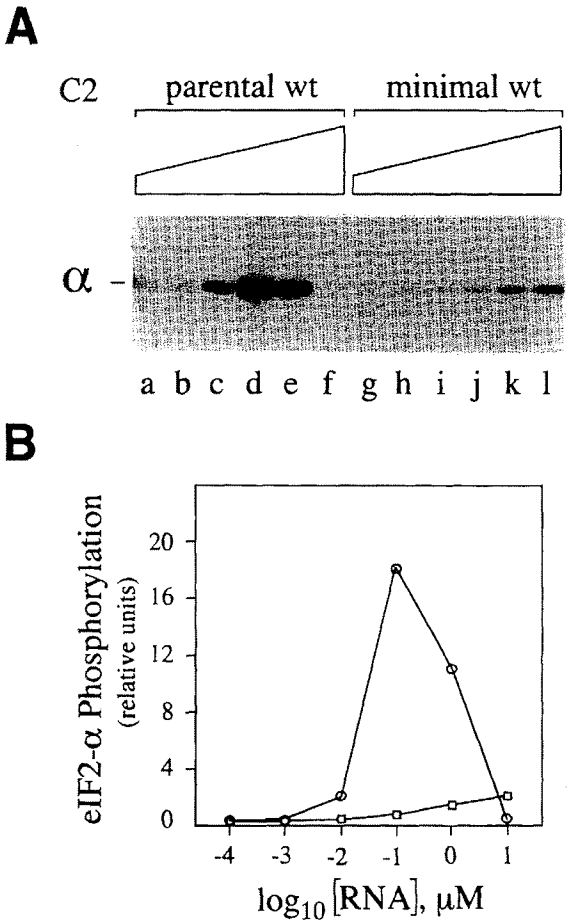


FIGURE 9: In vitro activation of PKR by minimal-length RNAs. (A) Shown is an autoradiogram of a 10% SDS−polyacrylamide gel corresponding to the region of the α subunit of eIF-2. Parental clone 2 wt RNA (lanes a−f) and minimal-length clone 2 wt RNA samples (lanes g−l) were compared for their ability to activate purified PKR kinase as measured by eIF-2α phosphorylation. RNA was added as indicated to the standard reaction mixture which contained [γ-^{32}P]ATP and eIF-2. Concentrations of RNAs used were lanes a and g, 0.0001 μM; lanes b and h, 0.001 μM; lanes c and i, 0.01 μM; lanes d and j, 0.1 μM; lanes e and k, 1.0 μM; lanes f and l, 10 μM RNA. α denotes the position of ^{32}P-labeled α-subunit of eIF-2. (B) Quantitation of autoradiogram shown in panel A (○), full-length clone 2 RNA, (□) minimal-length clone 2 RNA.

RNAs selected by binding to the truncated H_6Tp20 PKR protein that consists of residues 1−184 also were able to activate the wild-type 551 residue PKR protein purified from human cells (Figures 3 and 9). However, binding and activation were not equivalent. For example, even though the five RNAs of family 1 shared a 17-nt consensus sequence and displayed similar K_d values with H_6Tp20 protein, they differed significantly in their ability to activate PKR autophosphorylation and eIF-2 phosphorylation. These differences in enzyme activation observed for selected family 1 RNAs may relate to differences in their ability to induce the appropriate conformational change in the full-length protein required for enzymatic activity. While dimerization of full-length PKR can occur both in the presence and absence of RNA binding activity (*64*−*66*), the presence of RNA appears to mediate a conformational change in the catalytic domain

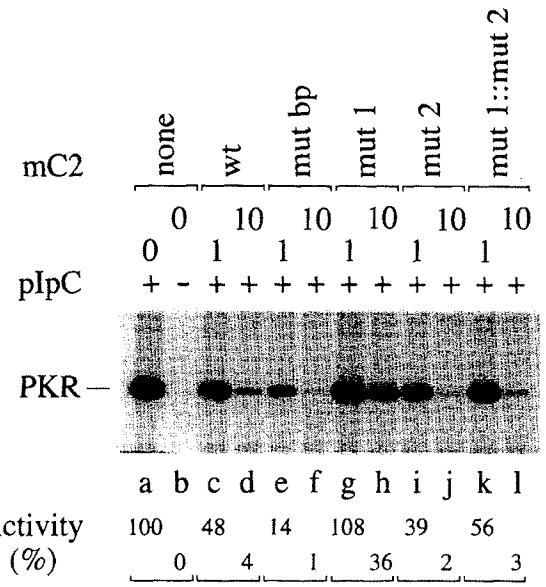


FIGURE 10: Antagonism of PKR activation by minimal-length RNAs. Shown is an autoradiogram of a 10% SDS−polyacrylamide gel. Minimal-length clone 2 (mC2) RNA either wt (lanes c and d) or mutant (lanes e−l), was examined for its ability to antagonize the activation of purified PKR kinase. PKR was incubated with selected RNAs at either 1.0 or 10 μM prior to the addition of 0.1 μg/mL (=0.8 nM) poly(rI)-poly(rC). Reactions carried out either in the absence of added RNA (lane b), or in the presence of the optimal concentration of poly(rI)-poly(rC) (lane a) but without added selected RNA, are shown as controls. PKR denotes the position of ^{32}P-labeled PKR. The autophosphorylation activity of PKR kinase was quantified and is expressed as a percent of the activity observed in the presence of an optimal concentration of poly(rI)-poly(rC).

of the PKR protein required for activation of kinase activity (*67*, *68*). Furthermore, both genetic (*69*) and biophysical (*68*) analyses suggest that the active form of PKR most likely is a dimer together with a single RNA molecule.

The double-mutant, base-pair mutant, and mutant 2 minimal-length clone 2 RNAs all antagonized the dsRNA-mediated activation of the full-length PKR kinase autophosphorylation as did the wt minimal-length clone 2 RNA (Figure 10). Somewhat surprisingly, however, binding of the mutant 2 RNA to H_6Tp20 protein was not detected by a native-gel mobility-shift assay, whereas the double-mutant, base-pair mutant and wt minimal-length clone 2 RNAs did bind (Figure 7B). This apparent contradiction may reflect the differing sensitivity of the two assays, or alternatively may indicate a transient RNA−protein interaction not detected by the gel-shift assay.

The dsRBD Can Tolerate Non-Watson−Crick Structures. Secondary structures for full-length RNAs from clones 2 and 11 were established by sequence alignment, free-energy minimization, structure mapping, boundary experiments, site-directed mutagenesis, and footprinting experiments. These data are most consistent with the secondary structures shown (Figure 4). These structures include a 3′-hairpin (P3/L3) with a stem that consists of 14 bp counting Watson−Crick, G−U wobble pairs, and the tandem A−G mismatch pair, and a shorter 5′-stem-loop (P2/L2) that is 8 bp in length and is interrupted by a C−U mismatch in one case.

Of these two stem-loop structures, the 3′-most stem loop appears to be the most important for protein binding as judged by the following group of experiments. Boundary experiments probing the 3′-end of the RNA reveal that only the four 3′-most single-stranded nucleotides can be removed from the RNA without affecting binding and that the 3′-stem itself is absolutely required for binding. Boundary experiments examining the 5′-end length requirements of the RNA show that while the 3′-stem-loop is not itself sufficient to bind H_6Tp20, the 5′-end requirements are poorly defined. Inspection of predicted foldings for various 5′-end truncations reveals that they can form minihelices 3−5 bp in length. It is plausible that a minihelix and the noncontiguous 14 bp 3′-stem-loop combine to provide the minimal-length binding site of 16 bp (*22*). This was an unexpected result and suggests that, in general, binding sites for PKR protein may be assembled from several noncontiguous helices. Recent secondary and tertiary structure models for VA I RNA depict two noncontiguous stem loops as being involved in binding PKR, termed the apical stem and stem 7 (*36*). Our results are consistent with such a binding site.

Additionally, for clone 2 RNA, the only observed footprint was in the 3′-stem-loop. For clone 11 RNA, the 3′-stem-loop was protected at 12 different positions while the 5′-stem-loop was protected at only two positions. Interestingly, five of the 12 3′-stem-loop protections occur in the loop itself with one enhancement adjacent to the loop (Figure 4). Protections of the loop may arise from occlusion of the loop by bound H_6Tp20, or perhaps binding of H_6Tp20 to the loop. Enhancement of RNase V1 cleavage may arise from conformational changes of the RNA upon protein binding that render the site optimal for cleavage. The pattern of protections and enhancement observed in the 3′-stem-loop of clone 11 RNA suggests that H_6Tp20 may interact differently with the loop-proximal and loop-distal portions of the stem.

An unexpected feature of the 3′-stem-loop is the presence of a non-Watson−Crick secondary structural element, the tandem A−G mismatch. The role of the tandem A−G mismatch in protein binding was investigated by construction of minimal RNAs containing the 3′-stem-loop plus two extra base pairs to provide a 16 bp binding site. Converting the tandem A−G mismatch to Watson−Crick base pairs had no significant effect on binding affinity, suggesting that the tandem A−G mismatch is merely tolerated within the flanking dsRNA stems. Its selective advantage may then be manifest at some level other than protein binding such as the reverse transcription or PCR amplification step (see below). A separate question is whether the tandem A−G mismatch makes specific contacts with the protein and confers site specificity, as suggested for other secondary structure defects in RNAs (*35*).

The tandem A−G mismatch has been investigated both functionally and structurally by Wu et al. (*56*, *70*). Tandem A−G mismatches constitute thermodynamically stable non-Watson−Crick motifs, stabilizing a helix by as much as −1.2 kcal/mol depending on closing base pairs (*70*). NMR investigations reveal that this stabilization arises because of a stacked, imino-hydrogen-bonded conformation with two hydrogen bonds per A−G mismatch (*56*). The tandem A−G mismatch perturbs the structure in only a minor fashion. The backbone is expanded at the mismatch and the helical rise is decreased; however, an overall A-form geometry is maintained. In addition, these minor changes are propagated only to the neighboring base pairs (*56*). Weak cleavage of the tandem A−G mismatch in clone 11 RNA by RNase V1 (Figures 4 and 5) is consistent with the stacked A−G structure and the known property of RNase V1 to cleave many non-Watson−Crick stacked structures including wobble G−U pairs (*60*). In contrast to the tandem A−G mismatch, tandem G−A mismatches can form a sheared conformation and introduce severe distortion into the backbone (*71*).

Several reports support the regulation of PKR by non-double-stranded RNAs. Early studies showed that PKR could not be activated if a poly(C):poly(I) duplex were interrupted with a G−I mismatch every 7 bp, on average, but could be partially or completely activated if the G−I mismatch were present only every 15 or 45 bp on average, respectively (*72*). PKR can also be activated by reovirus s1 mRNA (*25*) and the 3′-UTR from human α-tropomyosin mRNA (*33*) and inhibited by the adenovirus VA I RNA, Epstein−Barr virus EBER RNAs, and HIV-1 TAR RNA (*32*, *73*), all of which contain secondary structure defects. RNase T1 protection assays reveal that PKR can bind to the genomic RNA from hepatitis delta agent between nucleotides 710−872 (*35*). The δ agent genomic RNA does not contain a tandem A−G mismatch in this stretch of residues, but does contain an internal 5′AA3′/5′AG3′ mismatch (*35*). This latter mismatch is found in clone 14 RNA, the member of family 1 with an exception to the tandem A−G mismatch (Figure 2). It is thus possible that the two mismatch types form similar, dsRNA-like structures. Wu et al. (*70*) surveyed 24 small-subunit and 51 large-subunit prokaryotic and eukaryotic phylogenetic rRNA secondary structures for tandem mismatches. The tandem G−A and the 5′AA3′/5′GA3′ mismatches were found 71 and 54 times, respectively, while the thermodynamically stable tandem A−G mismatch and the 5′AA3′/5′AG3 were not found (*70*). One possible explanation for the absence of these mismatches is that they can regulate dsRBM-containing proteins and as such are not prevalent motifs. Double-stranded RNA deaminases contain multiple copies of the dsRBM and are able to convert A−U pairs to I−U pairs by a deamination reaction (*74*−*77*). Multiple A−U to I−U conversions can occur within the same helix (*27*). Inspection of the I−U pair geometry reveals that it can form a two imino-hydrogen-bonded wobble pair. Apparently, the dsRBM can also bind to double-stranded RNA containing a limited number of I−U wobble pairs.

Design of the experiments in this study may be biased against the selection of long, purely double-stranded RNA regions. Such regions require reverse transcriptase to read through highly structured RNAs. Although the reverse transcription step was performed at 60 °C, such structures were probably selected against. The goal of this study was identification of non-Watson−Crick motifs; therefore, bias against long, regular double-stranded RNAs may have aided the outcome.

Double-stranded regions were found to form between the random region and the fixed primer-binding sites, with the longest duplexes present at the 3′-end. Such pairing may be biased by experimental design. The 3′-primer-binding site is bound first by the bottom-strand primer during reverse transcription. Both primer-binding sites are bound during PCR. Primer may first bind to accessible single-stranded regions adjacent to the RNA stem loop and then invade and

disrupt the neighboring RNA stem loop affording replication. Such disruptions, readily achieved at the 3′-end of the molecule during reverse transcription, are not possible for stems formed within the randomized region only. Since not all non-Watson−Crick possibilities are present in the fixed-sequence primer-binding sites, an exhaustive study of the role of non-Watson−Crick structures in dsRBM binding has not been carried out. Nevertheless, several important conclusion can be made: (1) the binding site for the dsRBD has been found to be composed of RNAs that are similar to, or mimic, A-form dsRNA, one expected outcome based on numerous previous studies; (2) the tandem A−G mismatch can support protein binding; (3) noncontiguous helices can combine to form a site for protein binding; and (4) an experimental basis has been established which will aid design of further selection experiments to examine the role of RNA structure in dsRBD binding.

ACKNOWLEDGMENT

We thank Anne Gooding and Elaine Podell for oligonucleotide synthesis.

REFERENCES

1. Samuel, C. E. (1993) *J. Biol. Chem. 268*, 7603−7606.
2. Clemens, M. J. Protein Kinases That Phosphorylate eIF2 and eIF2B, and Their Role in Eukaryotic Cell Translational Control. in *Translational Control* (Hershey, J. W. B., Mathews, M. B., and Sonenberg, N., Eds.) pp 139−172, Cold Spring Harbor Laboratory, Plainview, NY.
3. Kumar, A., Yang, Y. L., Flati, V., Der, S., Kadereit, S., Deb, A., Haque, J., Reis, L., Weissmann, C., and Williams, B. R. (1997) *EMBO J. 16*, 406−416.
4. Wong, A. H., Tam, N. W., Yang, Y. L., Cuddihy, A. R., Li, S., Kirchhoff, S., Hauser, H., Decker, T., and Koromilas, A. E. (1997) *EMBO J. 16*, 1291−1304.
5. Roy, S., Katze, M. G., Parkin, N. T., Edery, I., Hovanessian, A. G., and Sonenberg, N. (1990) *Science 247*, 1216−1219.
6. Benkirane, M., Neuveut, C., Chun, R. F., Smith, S. M., Samuel, C. E., Gatignol, A., and Jeang, K. T. (1997) *EMBO J. 16*, 611−624.
7. Nagai, K., Wong, A. H., Li, S., Tam, W. N., Cuddihy, A. R., Sonenberg, N., Mathews, M. B., Hiscott, J., Wainberg, M. A., and Koromilas, A. E. (1997) *J. Virol. 71*, 1718−1725.
8. Kitajewski, J., Schneider, R. J., Safer, B., Munemitsu, S. M., Samuel, C. E., Thimmappaya, B., and Shenk, T. (1986) *Cell 45*, 195−200.
9. Mathews, M. B., and Shenk, T. (1991) *J. Virol. 65*, 5657−5662.
10. Lee, S. B., and Esteban, M. (1994) *Virology 199*, 491−496.
11. Der, S. D., Yang, Y. L., Weissmann, C., and Williams, B. R. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3279−3283.
12. Kibler, K. V., Shors, T., Perkins, K. B., Zeman, C. C., Banaszak, M. P., Biesterfeldt, J., Langland, J. O., and Jacobs, B. L. (1997) *J. Virol. 71*, 1992−2003.
13. Lee, S. B., Rodriguez, D., Rodriguez, J. R., and Esteban, M. (1997) *Virology 231*, 81−88.
14. Katze, M. G., Wambach, M., Wong, M.-L., Garfinkel, M., Meurs, E., Chong, K., Williams, B. R. G., Hovanessian, A. G., and Barber, G. N. (1991) *Mol. Cell. Biol. 11*, 5497−5505.
15. Feng, G.-S., Chong, K., Kumar, A., and Williams, B. R. G. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5447−5451.
16. Green, S. R., and Mathews, M. B. (1992) *Genes Dev. 6*, 2478−2490.
17. McCormack, S. J., Thomis, D. C., and Samuel, C. E. (1992) *Virology 188*, 47−56.
18. Patel, R. C., and Sen, G. C. (1992) *J. Biol. Chem. 267*, 7671−7676.
19. St Johnston, D., Brown, N. H., Gall, J. G., and Jantsch, M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10979−10983.
20. Kharrat, A., Macias, M. J., Gibson, T. J., Nilges, M., and Pastore, A. (1995) *EMBO J. 14*, 3572−3584.
21. Bycroft, M., Grünert, S., Murzin, A. G., Proctor, M., and St. Johnston, D. (1995) *EMBO J. 14*, 3563−3571.
22. Bevilacqua, P. C., and Cech, T. R. (1996) *Biochemistry 35*, 9983−9994.
23. Hunter, T., Hunt, T., Jackson, R. J., and Robertson, H. D. (1975) *J. Biol. Chem. 250*, 409−417.
24. Samuel, C. E. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 600−604.
25. Bischoff, J. R., and Samuel, C. E. (1989) *Virology 172*, 106−115.
26. Manche, L., Green, S. R., Schmedt, C., and Mathews, M. B. (1992) *Mol. Cell. Biol. 12*, 5238−5248.
27. Polson, A. G., and Bass, B. L. (1994) *EMBO J. 13*, 5701−5711.
28. Schweisguth, D. C., Chelladurai, B. S., Nicholson, A. W., and Moore, P. B. (1994) *Nucleic Acids Res. 22*, 604−612.
29. Katze, M. G. (1995) *Trends Microbiol. 3*, 75−78.
30. Proud, C. G. (1995) *Trends Biochem. Sci. 20*, 241−246.
31. Jacobs, B. L., and Langland, J. O. (1996) *Virology 219*, 339−349.
32. Robertson, H. D., and Mathews, M. B. (1996) *Biochimie 78*, 909−914.
33. Davis, S., and Watson, J. C. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 508−513.
34. Robertson, H. D., Manche, L., and Mathews, M. B. (1996) *J. Virol. 70*, 5611−5617.
35. Circle, D. A., Neel, O. D., Robertson, H. D., Clarke, P. A., and Mathews, M. B. (1997) *RNA 3*, 438−448.
36. Ma, Y., and Mathews, M. B. (1996) *RNA 2*, 937−951.
37. Clarke, P. A., Sharp, N. A., and Clemens, M. J. (1990) *Eur. J. Biochem. 193*, 635−641.
38. Sharp, T. V., Schwemmle, M., Jeffrey, I., Laing, K., Mellor, H., Proud, C. G., Hilse, K., and Clemens, M. J. (1993) *Nucleic Acids Res 21*, 4483−4490.
39. Gunnery, S., Rice, A. P., Robertson, H. D., and Mathews, M. B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8687−8691.
40. Roy, S., Agy, M., Hovanessian, A. G., Sonenberg, N., and Katze, M. (1991) *J. Virol. 65*, 632−640.
41. Maitra, R. K., McMillan, N. A. J., Desai, S., McSwiggen, J., Hovanessian, A. G., Sen, G., Williams, B. R. G., and Silverman, R. H. (1994) *Virology 204*, 823−827.
42. Zhu, S., Romano, P. R., and Wek, R. C. (1997) *J. Biol. Chem. 272*, 14434−14441.
43. Thomis, D. C., Doohan, J. P., and Samuel, C. E. (1992) *Virology 188*, 33−46.
44. Tuerk, C., and Gold, L. (1990) *Science 249*, 505−510.
45. Bartel, D. P., and Szostak, J. W. (1993) *Science 261*, 1411−1418.
46. Czerny, T. (1996) *Nucleic Acids Res. 24*, 985−986.
47. Milligan, J. F., and Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51−62.
48. Schmedt, C., Green, S. R., Manche, L., Taylor, D. R., Ma, Y., and Mathews, M. B. (1995) *J. Mol. Biol. 249*, 29−44.
49. Zaug, A. J., Dávila-Aponte, J. A., and Cech, T. R. (1994) *Biochemistry 33*, 14935−14947.
50. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995) *Nucleic Acids Res. 23*, 2677−2684.
51. Vlassov, V. V., Zuber, G., Felden, B., Behr, J.-P., and Giege, R. (1995) *Nucleic Acids Res. 23*, 3161−3167.
52. Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527−2538.
53. Samuel, C. E., Knutson, G. S., Berry, M. J., Atwater, J. A., and Lasky, S. R. (1986) *Methods Enzymol. 119*, 499−516.
54. Thomis, D. C., and Samuel, C. E. (1995) *J. Virol. 69*, 5195−5198.
55. Doudna, J. A., Cech, T. R., and Sullenger, B. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 2355−2359.
56. Wu, M., SantaLucia, J., Jr., and Turner, D. H. (1997) *Biochemistry 36*, 4449−4460.

57. Jaeger, J. A., Turner, D. H., and Zuker, M. (1990) *Methods Enzymol. 183*, 281−306.
58. Serra, M. J., and Turner, D. H. (1995) *Methods Enzymol. 259*, 242−261.
59. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P., and Ehresmann, B. (1987) *Nucleic Acids Res. 15*, 9109−9128.
60. Lockard, R. E., and Kumar, A. (1981) *Nucleic Acids Res. 9*, 5125−5140.
61. Kolchanov, N. A., Titov, I. I., Vlassova, I. E., and Vlassov, V. V. (1996) *Prog. Nucleic Acid Res. Mol. Biol. 53*, 131−196.
62. Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., and Turner, D. (1985) *Biochemistry 24*, 4533−4539.
63. Dunn, J. J., and Studier, F. W. (1975) *J. Mol. Biol. 99*, 487−499.
64. Patel, R. C., Stanton, P., McMillan, N. M. J., Williams, B. R. G., and Sen, G. C. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 8283−8287.
65. Ortega, L. G., McCotter, M. D., Henry, G. L., McCormack, S. J., Thomis, D. C., and Samuel, C. E. (1996) *Virology 215*, 31−39.
66. Wu, S., and Kaufman, R. J. (1997) *J. Biol. Chem. 272*, 1291−1296.
67. Bischoff, J. R., and Samuel, C. E. (1985) *J. Biol. Chem. 260*, 8237−8239.
68. Carpick, B. W., Graziano, V., Schneider, D., Maitra, R. K., Lee, X., and Williams, B. R. G. (1997) *J. Biol. Chem. 272*, 9510−9516.
69. Romano, P. R., Green, S. R., Barber, G. N., Mathews, M. B., and Hinnebusch, A. G. (1995) *Mol. Cell. Biol. 15*, 365−378.
70. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
71. SantaLucia, J., Jr., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
72. Minks, M. A., West, D. K., Benvin, S., and Baglioni, C. (1979) *J. Biol. Chem. 254*, 10180−10183.
73. Clemens, M. J., Laing, K. G., Jeffrey, I. W., Schofield, A., Sharp, T. V., Elia, A., Matys, V., James, M. C., and Tilleray, V. J. (1994) *Biochimie 76*, 770−778.
74. Kim, U., Wang, Y., Sanford, T., Zeng, Y., and Nishikura, K. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11457−11461.
75. O'Connell, M. A., Krause, S., Higuchi, M., Hsuan, J. J., Totty, N. F. T., Jenny, A., and Keller, W. (1995) *Mol. Cell. Biol. 15*, 1389−1397.
76. Patterson, J. B., and Samuel, C. E. (1995) *Mol. Cell. Biol. 15*, 5376−5388.
77. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P. H., and Higuchi, M. (1996) *Nature 379*, 460−464.

BI980113J

RNA interference is mediated by 21- and 22-nucleotide RNAs

Sayda M. Elbashir, Winfried Lendeckel, and Thomas Tuschl[1]

Department of Cellular Biochemistry, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

Double-stranded RNA (dsRNA) induces sequence-specific posttranscriptional gene silencing in many organisms by a process known as RNA interference (RNAi). Using a *Drosophila* in vitro system, we demonstrate that 21- and 22-nt RNA fragments are the sequence-specific mediators of RNAi. The short interfering RNAs (siRNAs) are generated by an RNase III–like processing reaction from long dsRNA. Chemically synthesized siRNA duplexes with overhanging 3′ ends mediate efficient target RNA cleavage in the lysate, and the cleavage site is located near the center of the region spanned by the guiding siRNA. Furthermore, we provide evidence that the direction of dsRNA processing determines whether sense or antisense target RNA can be cleaved by the siRNA–protein complex.

[*Key Words*: RNAi; posttranscriptional gene silencing; dsRNA; siRNA]

Received October 25, 2000; revised version accepted November 28, 2000.

The term RNA interference (RNAi) was coined after the discovery that injection of dsRNA into the nematode *Caenorhabditis elegans* leads to specific silencing of genes highly homologous in sequence to the delivered dsRNA (Fire et al. 1998). RNAi was also observed subsequently in insects (Kennerdell and Carthew 1998), frog (Oelgeschlager et al. 2000), and other animals including mice (Svoboda et al. 2000; Wianny and Zernicka-Goetz 2000) and is likely to also exist in human. RNAi is closely linked to the posttranscriptional gene-silencing (PTGS) mechanism of cosuppression in plants and quelling in fungi (Cogoni and Macino 1999; Catalanotto et al. 2000; Dalmay et al. 2000; Ketting and Plasterk 2000; Mourrain et al. 2000; Smardon et al. 2000), and some components of the RNAi machinery are also necessary for posttranscriptional silencing by cosuppression (Catalanotto et al. 2000; Dernburg et al. 2000; Ketting and Plasterk 2000). The topic has been reviewed recently (Fire 1999; Sharp 1999; Bass 2000; Bosher and Labouesse 2000; Plasterk and Ketting 2000; Sijen and Kooter 2000; see also the entire issue of *Plant Molecular Biology*, Vol. 43, issue 2/3, 2000).

The natural function of RNAi and cosuppression appears to be protection of the genome against invasion by mobile genetic elements such as transposons and viruses, which produce aberrant RNA or dsRNA in the host cell when they become active (Jensen et al. 1999; Ketting et al. 1999; Ratcliff et al. 1999; Tabara et al. 1999; Malinsky et al. 2000). Specific mRNA degradation prevents transposon and virus replication, although some viruses are able to overcome or prevent this process by expressing proteins that suppress PTGS (Anandalakshmi et al. 2000; Lucy et al. 2000; Voinnet et al. 2000).

DsRNA triggers the specific degradation of homologous RNAs only within the region of identity with the dsRNA (Zamore et al. 2000). The dsRNA is processed to 21–23-nt RNA fragments (Zamore et al. 2000). These short fragments were also detected in extracts prepared from *Drosophila melanogaster* Schneider 2 cells that were transfected with dsRNA before cell lysis (Hammond et al. 2000) or after injection of radiolabeled dsRNA into *D. melanogaster* embryos (Yang et al. 2000) or *C. elegans* adults (Parrish et al. 2000). RNA molecules of similar size also accumulate in plant tissue that exhibits PTGS (Hamilton and Baulcombe 1999). It has been suggested that the 21–23-nt fragments are the guide RNAs for target recognition (Hamilton and Baulcombe 1999; Hammond et al. 2000), which is supported by the finding that the target mRNA is cleaved in 21–23-nt intervals (Zamore et al. 2000).

Here, we use the established *Drosophila* in vitro system (Tuschl et al. 1999; Zamore et al. 2000) to explore further the mechanism of RNAi. It is demonstrated that synthetic 21- and 22-nt RNAs, when base paired with 3′ overhanging ends, act as the guide RNAs for sequence-specific mRNA degradation. Short 30-bp dsRNAs are inefficiently processed to 21- and 22-nt RNAs, which may explain why they are ineffective in mediating RNAi. Furthermore, we define the target RNA cleavage sites relative to the 21- and 22-nt short interfering RNAs (siRNAs) and provide evidence that the direction of dsRNA processing determines whether a sense or an antisense

[1]Corresponding author.
E-MAIL ttuschl@mpibpc.gwdg.de; FAX 49-551-201-1197.
Article and publication are at www.genesdev.org/cgi/doi/10.1101/gad.862301.

target RNA can be cleaved by the siRNP endonuclease complex.

Length requirements for processing of dsRNA to 21- and 22-nt RNA fragments

Lysate prepared from *D. melanogaster* syncytial embryos recapitulates RNAi in vitro, providing a tool for biochemical analysis of the mechanism of RNAi (Tuschl et al. 1999; Zamore et al. 2000). In vitro and in vivo analysis of the length requirements of dsRNA for RNAi has revealed that short dsRNA (<150 bp) are less effective than longer dsRNAs in degrading target mRNA (Ngo et al. 1998; Tuschl et al. 1999; Caplen et al. 2000; Hammond et al. 2000). The reasons for reduction in mRNA degrading efficiency are not understood. We therefore examined the precise length requirement of dsRNA for target RNA degradation under optimized conditions in the *Drosophila* lysate. Three series of dsRNAs were synthesized and directed against firefly luciferase (*Pp*-luc) reporter RNA. The dual luciferase assay was used to monitor specific suppression of target RNA expression (Tuschl et al. 1999; (Fig. 1A,B). Specific inhibition of target RNA expression was detected for dsRNAs as short as 38 bp, but dsRNAs of 29–36 bp were not effective in this process. The effect was independent of the target position and the degree of inhibition of *Pp*-luc mRNA expression correlated with the length of the dsRNA; that is, long dsRNAs were more effective than short dsRNAs.

It has been suggested that the 21–23-nt RNA fragments generated by processing of dsRNAs are the mediators of RNA interference and cosuppression (Hamilton and Baulcombe 1999; Hammond et al. 2000; Zamore et al. 2000). We therefore analyzed the rate of 21–23-nt fragment formation for a subset of dsRNAs ranging in size from 501 to 29 bp. Formation of 21–23-nt fragments in *Drosophila* lysate (Fig. 2) was readily detectable for 39–501 bp dsRNAs but was significantly delayed for the 29-bp dsRNA. This observation is consistent with a role of 21–23-nt fragments in guiding mRNA cleavage and provides an explanation for the lack of RNAi by 30-bp dsRNAs. The length dependence of 21–23 mer formation is likely to reflect a mechanism to prevent the undesired activation of RNAi by short intramolecular base-paired structures of cellular RNAs.

Mapping of the cleavage sites on sense and antisense target RNAs

Addition of dsRNA and 5′-capped target RNA to the *Drosophila* lysate results in sequence-specific degradation of the target RNA (Tuschl et al. 1999). The target mRNA is only cleaved within the region of identity with the dsRNA, and many of the target cleavage sites are separated by 21–23 nt (Zamore et al. 2000). Thus, the number of cleavage sites for a given dsRNA was expected to roughly correspond to the length of the dsRNA divided by 21. We mapped the target cleavage sites on a sense and an antisense target RNA that was 5′ radiolabeled at the cap (Zamore et al. 2000; Fig. 3A,B). Stable 5′ cleavage products were separated on a sequencing gel, and the position of cleavage was determined by comparison with a partial RNase T1 and an alkaline hydrolysis ladder from the target RNA.

Consistent with the previous observation (Zamore et al. 2000), all target RNA cleavage sites were located within the region of identity to the dsRNA. The 39-bp dsRNA produced a strong and a weak (often hardly de-

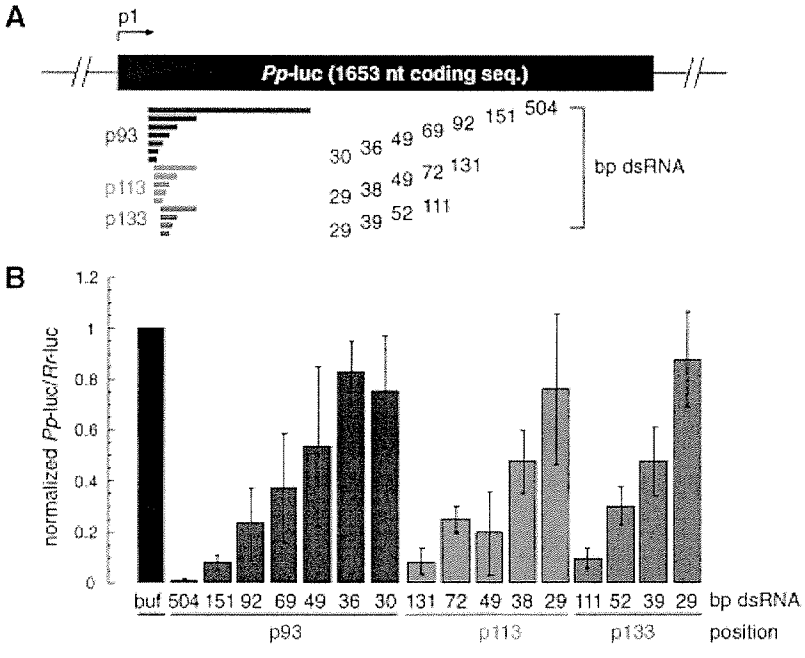


Figure 1. Double-stranded RNA as short as 38 bp can mediate RNAi. (*A*) Graphic representation of dsRNAs used for targeting *Pp*-luc mRNA. Three series of blunt-ended dsRNAs covering a range of 29–504 bp were prepared. The position of the first nucleotide of the sense strand of the dsRNA is indicated relative to the start codon of *Pp*-luc mRNA (p1). (*B*) RNA interference assay (Tuschl et al. 1999). Ratios of target *Pp*-luc to control *Rr*-luc activity were normalized to a buffer control (black bar). DsRNAs (5 nM) were preincubated in *Drosophila* lysate at 25°C for 15 min before the addition of 7-methyl-guanosine-capped *Pp*-luc and *Rr*-luc mRNAs (~50 pM). The incubation was continued for another hour and then analyzed by the dual luciferase assay (Promega). The data are the average from at least four independent experiments ±S.D.

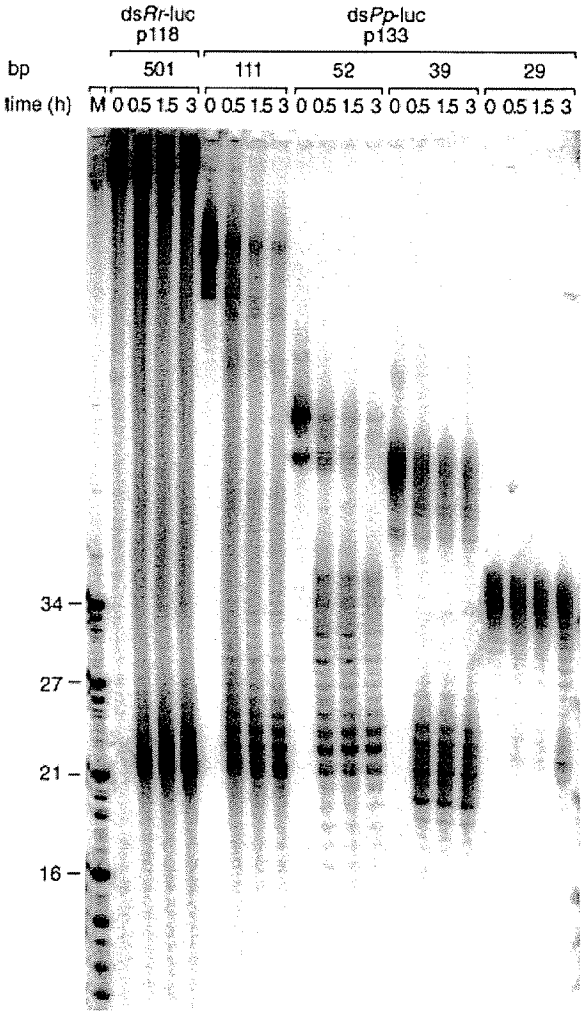


Figure 2. A 29-bp dsRNA is only slowly processed to 21–23-nt fragments. Time course of 21–23-mer formation from processing of internally ^{32}P-labeled dsRNAs (5 nM) in the *Drosophila* lysate. The length and source of the dsRNA are indicated. An RNA size marker (M) has been loaded in the left lane, and the fragment sizes are indicated. Double bands at time zero are caused by incompletely denatured dsRNA.

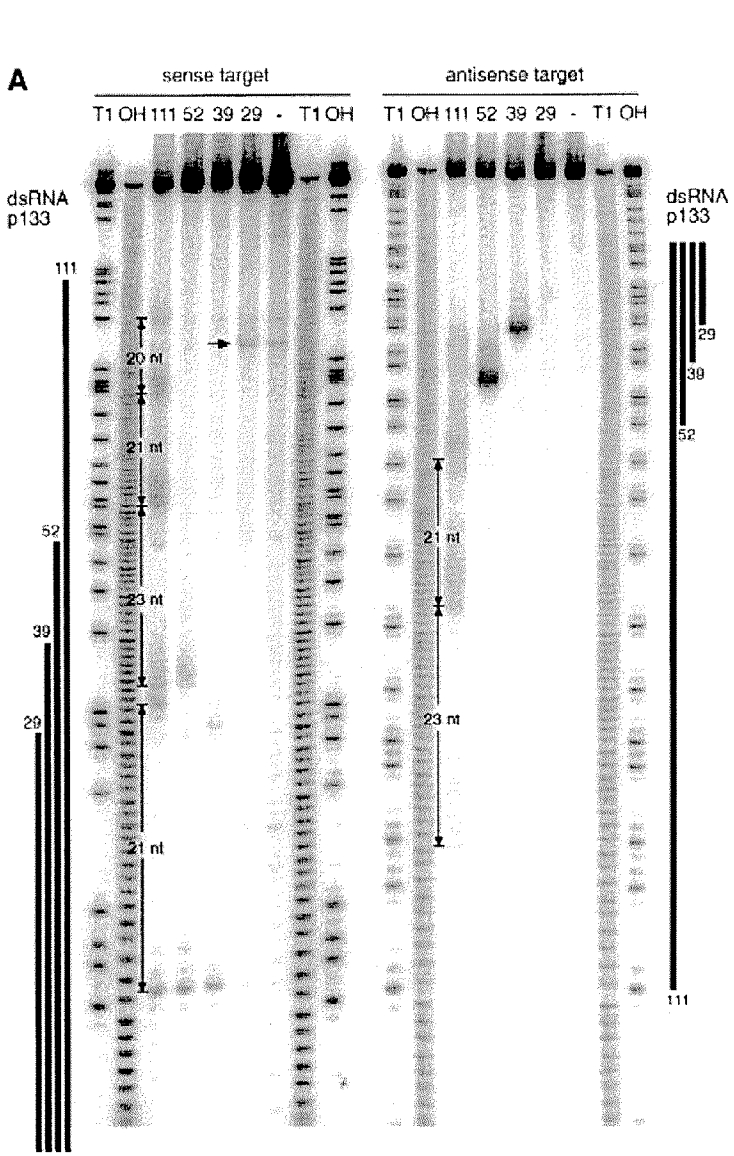
tectable) cleavage site in the sense target RNA separated by 19 nt. The antisense target was only cleaved once, by the 39-bp dsRNA. The predominant cleavage site of the sense strand and the cleavage site of the antisense strand are located 10 nt from the 5′ end of the region covered by the dsRNA (Fig. 3B). The 52-bp dsRNA, which shares the same 5′ end as the 39-bp dsRNA, produces the same strong cleavage site on the sense target, located 10 nt from the 5′ end of the region of identity with the dsRNA in addition to two weaker cleavage sites 23 and 24 nt downstream of the first site. The antisense target was only cleaved once, again 10 nt from the 5′ end of the region covered by its respective dsRNA. Mapping of the cleavage sites for the 38–49-bp dsRNAs shown in Figure 1 revealed that the first and predominant cleavage site was always located 7–10 nt downstream from the 5′ boundary of the region covered by the dsRNA (data not shown). This suggests that the point-of-target RNA cleavage can be determined by the end of the dsRNA and could imply that processing to 21–23mers starts from the ends of the duplex.

Cleavage sites on sense and antisense targets for the longer 111-bp dsRNA were much more frequent than anticipated, and most of them appear in clusters separated by 20–23 nt (Fig. 3A,B). As for the shorter dsRNAs, the first cleavage site on the sense target is 10 nt from the 5′ end of the region spanned by the dsRNA, and the first cleavage site on the antisense target is located 9 nt from the 5′ end of region covered by the dsRNA. It is unclear what causes this disordered cleavage, but one possibility could be that longer dsRNAs may not only get processed from the ends but also internally, or there are some specificity determinants for dsRNA processing that we do not yet understand. Some irregularities to the 21–23 nt spacing were also noted previously (Zamore et al. 2000).

dsRNA is processed to 21- and 22-nt RNAs by an RNase III–like mechanism

To understand better the molecular basis of dsRNA processing and target RNA recognition, we decided to analyze the sequences of the 21–23-nt fragments generated by processing of 39-, 52-, and 111-bp dsRNAs in the *Drosophila* lysate. We first examined the 5′ and 3′ termini of the RNA fragments. Periodate oxidation of gel-purified 21–23-nt RNAs followed by β-elimination indicated the presence of a terminal 2′ and 3′ hydroxyl (data not shown). The 21–23mers were also responsive to alkaline phosphatase treatment, implying the presence of a 5′ terminal phosphate (data not shown). The presence of 5′ phosphate and 3′ hydroxyl termini suggests that the dsRNA could be processed by an enzymatic activity similar to *Escherichia coli* RNase III (for reviews, see Dunn 1982; Nicholson 1999; Robertson 1982, 1990).

To directionally clone the 21–23-nt RNA fragments, 3′ and 5′ adapter oligonucleotides were ligated to the purified 21–23 mers using T4 RNA ligase. The ligation products were reverse transcribed, PCR-amplified, concatamerized, cloned, and sequenced. Over 220 short RNAs were sequenced from dsRNA processing reactions of the 39-, 52-, and 111-bp dsRNAs (Fig. 4A). We found the following length distribution: 1% 18 nt, 5% 19 nt, 12% 20 nt, 45% 21 nt, 28% 22 nt, 6% 23 nt, and 2% 24 nt. Sequence analysis of the 5′ terminal nucleotide of the processed fragments indicated that oligonucleotides with a 5′ guanosine were underrepresented. This bias was most likely introduced by T4 RNA ligase, which discriminates against 5′ phosphorylated guanosine as donor oligonucleotide (Romaniuk et al. 1982); no significant sequence bias was seen at the 3′ end. Many of the ~21-nt fragments originating from the 3′ ends of the sense or antisense strand of the duplexes include 3′ nucleotides that are derived from untemplated addition of nucleotides during RNA synthesis using T7 RNA



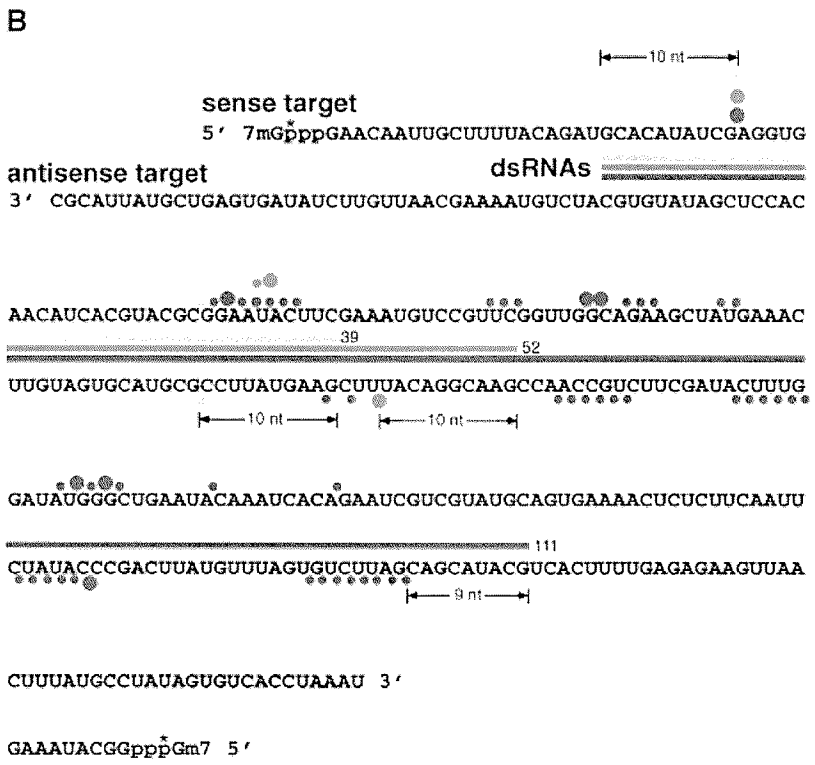


Figure 3. Mapping of sense and antisense target RNA cleavage sites. (*A*) Denaturing gel electrophoresis of the stable 5′ cleavage products produced by 1 h incubation of 10 nM sense or antisense RNA ^{32}P-labeled at the cap with 10 nM dsRNAs of the p133 series in *Drosophila* lysate. Length markers were generated by partial nuclease T1 digestion and partial alkaline hydrolysis (OH) of the cap-labeled target RNA. The regions targeted by the dsRNAs are indicated as black bars on both sides. The 20–23-nt spacing between the predominant cleavage sites for the 111-bp dsRNA is shown. The horizontal arrow indicates unspecific cleavage not caused by RNAi. (*B*) Position of the cleavage sites on sense and antisense target RNAs. The sequences of the capped 177-nt sense and 180-nt antisense target RNAs are represented in antiparallel orientation such that complementary sequence are opposing each other. The region targeted by the different dsRNAs are indicate by differently colored bars positioned between sense and antisense target sequences. Cleavage sites are indicated by circles (large circle for strong cleavage, small circle for weak cleavage). The ^{32}P-radiolabeled phosphate group is marked by an asterisk.

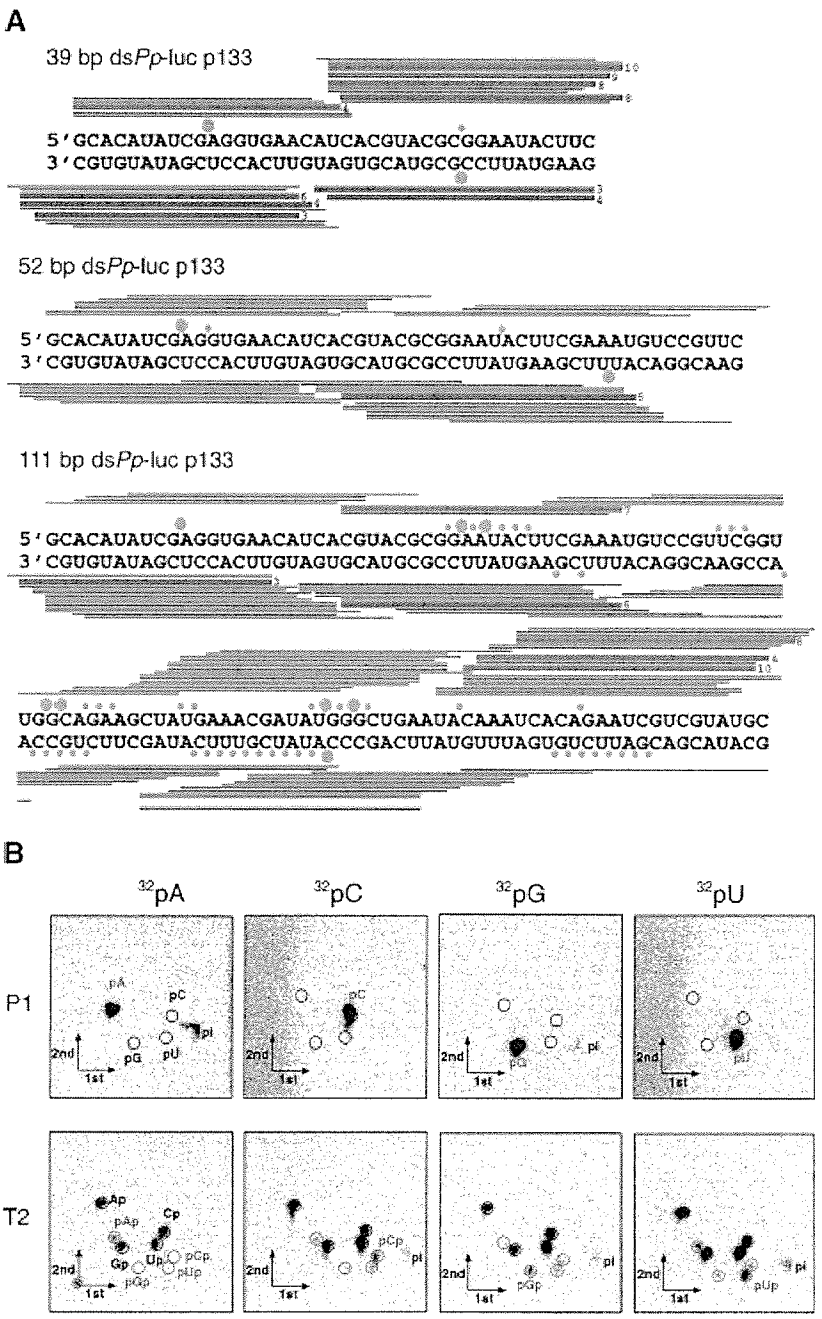


Figure 4. 21- and 22-nt RNA fragments are generated by an RNase III–like mechanism. (*A*) Sequences of ~21-nt RNAs after dsRNA processing. The ~21-nt RNA fragments generated by dsRNA processing were directionally cloned and sequenced. Oligoribonucleotides originating from the sense strand of the dsRNA are indicated as blue lines; those originating from the antisense strand are red lines. Thick bars are used if the same sequence was present in multiple clones, the number at the right indicating the frequency. The target RNA cleavage sites mediated by the dsRNA are indicated as orange circles: large circle for strong cleavage, small circle for weak cleavage (see Fig. 3B). Circles on top of the sense strand indicated cleavage sites within the sense target, and circles at the bottom of the dsRNA indicate cleavage site in the antisense target. Up to five additional nucleotides were identified in ~21-nt fragments derived from the 3′ ends of the dsRNA. These nucleotides are random combinations of predominantly C, G, or A residues and were most likely added in an untemplated fashion during T7 transcription of the dsRNA-constituting strands. (*B*) Two-dimensional TLC analysis of the nucleotide composition of ~21-nt RNAs. The ~21-nt RNAs were generated by incubation of internally radiolabeled 504-bp *Pp*-luc dsRNA in *Drosophila* lysate, gel purified, and then digested to mononucleotides with nuclease P1 (top row) or ribonuclease T2 (bottom row). The dsRNA was internally radiolabeled by transcription in the presence of one of the indicated α-^{32}P nucleoside triphosphates. Radioactivity was detected by phosphorimaging. Nucleoside 5′-monophosphates, nucleoside 3′-monophosphates, nucleoside 5′,3′-diphosphates, and inorganic phosphate are indicated as pN, Np, pNp, and p_i, respectively. Black circles indicate UV-absorbing spots from nonradioactive carrier nucleotides. The 3′,5′-diphosphates (red circles) were identified by comigration with radiolabeled standards prepared by 5′-phosphorylation of nucleoside 3′-monophosphates with T4 polynucleotide kinase and γ-^{32}P-ATP (data not shown).

polymerase. Interestingly, a significant number of endogenous *Drosophila* ~21-nt RNAs were also cloned, some of them from LTR and non-LTR retrotransposons (data not shown). This is consistent with a possible role for RNAi in transposon silencing (Ketting et al. 1999; Tabara et al. 1999).

The ~21-nt RNAs appear in clustered groups (Fig. 4A) that cover the entire dsRNA sequences. For the 39-bp dsRNA, two clusters of ~21-nt RNAs were found from each dsRNA-constituting strand (including overhanging 3′ ends). Only one of the clusters from each strand can be correlated with a strong cleavage hot spot on the target sense or antisense RNA (Fig. 3A,B), indicating that dsRNA processing produced primarily two functional small RNAs originating from the 3′ ends of the duplex. Perhaps the ~21-nt RNAs are present in double-stranded form in the endonuclease complex, but only one of the strands can be used for target RNA recognition and cleavage.

The ~21-mer clusters for the 52- and 111-bp dsRNA

are less well defined when compared to the 39-bp dsRNA. The clusters are spread over regions of 25–30 nt most likely representing several distinct subpopulations of ~21-nt duplexes and, therefore, guiding target cleavage at several nearby sites. These cleavage regions are still predominantly separated by 20–23-nt intervals. The rules determining how dsRNA can be processed to ~21-nt fragments are not yet understood, but it was observed previously that the ~21–23-nt spacing of cleavage sites could be altered by a run of uridines (Zamore et al. 2000). The specificity of dsRNA cleavage by *E. coli* RNase III appears to be mainly controlled by antideterminants, that is, excluding some specific base pairs at given positions relative to the cleavage site (Zhang and Nicholson 1997). The sequence dependence of dsRNA processing and target RNA cleavage in RNAi needs to be examined further.

To test whether sugar-, base-, or cap-modification were present in processed ~21-nt RNA fragments, we incubated radiolabeled 505-bp *Pp*-luc dsRNA in lysate for 1 h, isolated the ~21-nt products, and digested it with P1 or T2 nuclease to mononucleotides. The nucleotide mixture was then analyzed by two-dimensional thin-layer chromatography (Fig. 4B). None of the four ribonucleotides were modified, as indicated by P1 or T2 digestion. We have previously analyzed adenosine to inosine conversion in the ~21-nt fragments (after a 2 h incubation) and detected a small extent (<0.7%) deamination (Zamore et al. 2000); shorter incubation in lysate (1 h) reduced this inosine fraction to barely detectable levels. RNase T2, which cleaves 3′ of the phosphodiester linkage, produced nucleoside 3′-phosphate and nucleoside 3′,5′-diphosphate, thereby indicating the presence of a 5′-terminal monophosphate. All four nucleoside 3′,5′-diphosphates were detected and indicate that the internucleotidic linkage was cleaved with little or no sequence specificity for the residue 3′ to the cleavage site, and according to the sequence analysis of the cloned ~21-nt fragments, no significant sequence bias was observed for the residue 5′ of the cleavage site. In summary, the ~21-nt fragments are unmodified and were generated from dsRNA such that 5′-monophosphates and 3′-hydroxyls were present at the 5′-ends. Analysis of the products of dsRNA processing indicated that the ~21-nt fragments are generated by a reaction with all the characteristics of an RNase III cleavage reaction (Dunn 1982; Robertson 1982, 1990; Nicholson 1999).

Synthetic 21- and 22-nt RNAs mediate target RNA cleavage

We chemically synthesized 21- and 22-nt RNAs, identical in sequence to some of the cloned ~21-nt fragments, and tested them for their ability to mediate target RNA degradation (Fig. 5A–C). The 21- and 22-nt RNA duplexes were incubated at 100 nM concentrations in the lysate, a 10- to 20-fold higher concentration than the 52-bp control dsRNA. Under these conditions, target RNA cleavage was readily detectable. Tenfold reduced concentrations of 21- and 22-nt duplexes (10 nM) still caused target RNA cleavage but to a smaller extent (data not shown). Increasing the duplex concentration from 100 to 1000 nM, however, did not further increase target degradation (data not shown), perhaps because of a limiting protein factor within the lysate. Single-stranded sense or antisense 21- and 22-nt RNAs at 100 nM concentration did not affect target RNA expression, most likely because single-stranded RNAs are not stable in the lysate and degraded to mononucleotides within minutes (data not shown). We also found that preannealing of the short antisense RNAs to the target mRNA before the addition of lysate had no effect on target RNA expression (data not shown).

RNase III makes two staggered cuts in both strands of the dsRNA, leaving a 3′ overhang of 2 nt. The 21- and 22-nt RNA duplexes with 2- or 3-nt overhanging 3′ ends (duplexes 1, 4, 6) were more efficient in reducing the target RNA expression than the corresponding blunt-ended dsRNAs (duplexes 2, 5, 7) or the dsRNA with 4 nt overhang (duplex 3). Duplexes 6 and 7 are generally more effective for RNAi than duplexes 1–5, probably as a consequence of target RNA accessibility (because of RNA self-structure or RNA-coating proteins) or because of sequence-specific effects in the reconstitution of the RNA-degrading complexes. The interference effects determined in the translation-based assay (Fig. 5B) correlate well with the intensity of the cleavage bands observed by targeting 5′ radiolabeled model substrates with the 21- and 22-nt RNA duplexes (Fig. 5C). Together, these data suggest that 2–3 nt of overhanging 3′ ends are beneficial for reconstitution of the RNAi nuclease complex and may be required for high-affinity binding of the short RNA duplex to the protein components. A 5′ terminal phosphate, although present after dsRNA processing, was not required to mediate target RNA cleavage and was absent from the short synthetic RNAs.

The synthetic 21- and 22-nt duplexes guided cleavage of sense as well as antisense targets within the region covered by the short duplex. This is interesting, considering that the presumably base-paired clusters of ~21-nt fragments derived from the 39-bp dsRNA (Fig. 2) can only be correlated to a predominant cleavage site on either the sense or the antisense target but not both. We interpret this result by suggesting that only one of two strands present in the ~21-nt duplex is able to guide target RNA cleavage and that the orientation of the ~21-nt duplex in the nuclease complex is determined by the initial direction of dsRNA processing. It also implies that the processed short RNAs are present in a tight ribonucleoprotein complex and do not dissociate and rebind during the time scale of the experiment. The presentation of an already perfectly processed ~21-nt duplex to the in vitro system, however, does allow formation of the active sequence-specific nuclease complex with two possible orientations of the symmetric RNA duplex. This results in cleavage of sense as well as antisense targets within the region of identity with the ~21-nt RNA duplex.

The target cleavage site is located near center of the region covered by the 21- or 22-nt RNAs, 11 or 12 nt

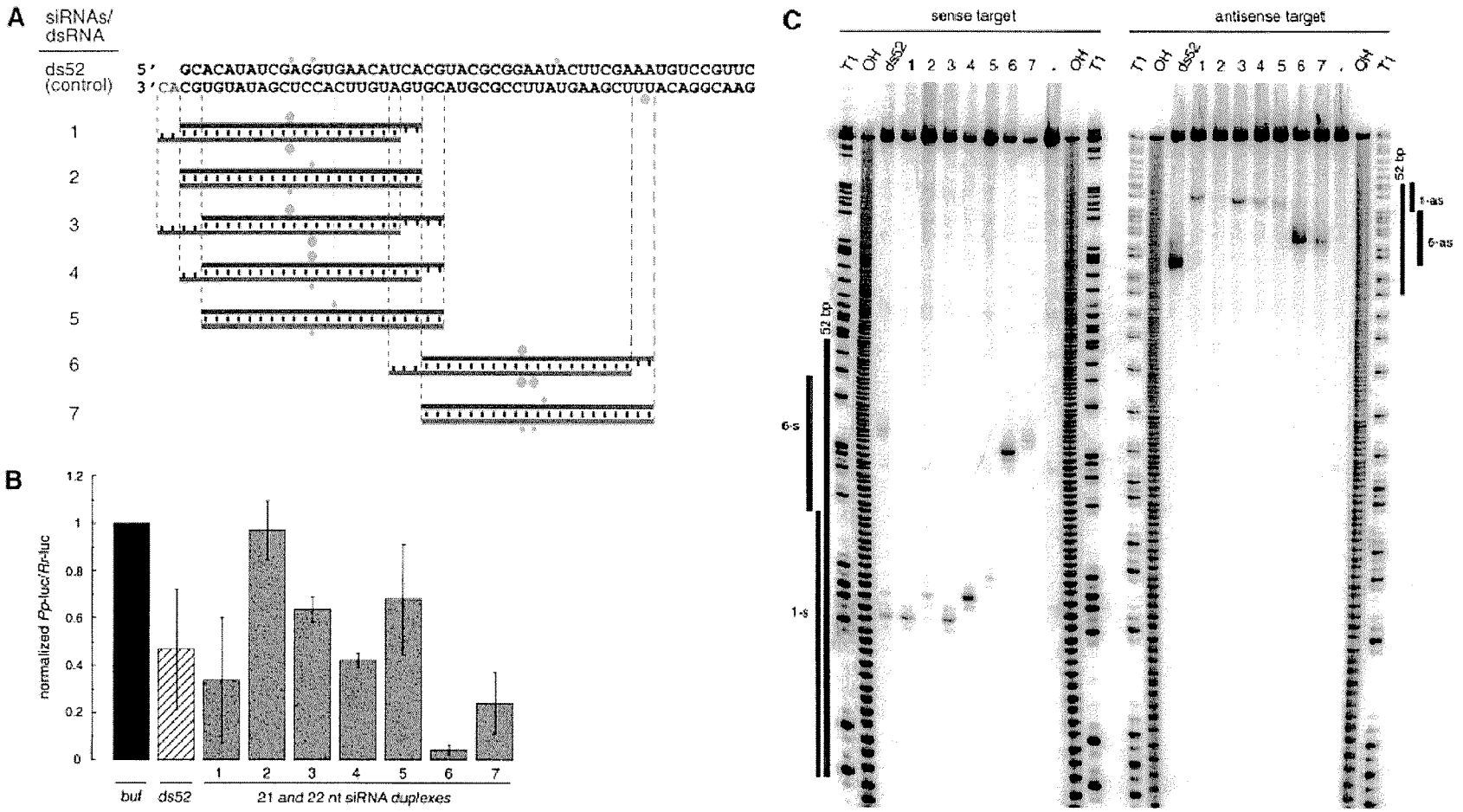


Figure 5. Synthetic 21- and 22-nt RNAs mediate target RNA cleavage. (*A*) Graphic representation of control 52-bp dsRNA and synthetic 21- and 22-nt dsRNAs. The sense strand of 21- and 22-nt short interfering RNAs (siRNAs) is shown in blue, the antisense strand in red. The sequences of the siRNAs were derived from the cloned fragments of 52- and 111-bp dsRNAs (Fig. 4A), except for the 22-nt antisense strand of duplex 5. The siRNAs in duplexex 6 and 7 were unique to the 111-bp dsRNA-processing reaction. The two 3′ overhanging nucleotides indicated in green are present in the sequence of the synthetic antisense strand of duplexes 1 and 3. Both strands of the control 52-bp dsRNA were prepared by in vitro transcription, and a fraction of transcripts may contain untemplated 3′ nucleotide addition. The target RNA cleavage sites directed by the siRNA duplexes are indicated as orange circles (see legend to Fig. 4A) and were determined as shown in Figure 5B. (*B*) RNA interference assay. To evaluate the efficiency of target RNA degradation, control 52-bp dsRNA (5 nM) or 21- and 22-nt RNA duplexes 1–7 (100 nM) targeting full-length *Pp*-luc mRNA were tested in the translation-based RNAi assay as described in Figure 1B. The relative luminescence of target to control luciferase normalized to a buffer control (buf) is blotted; error bars indicate standard deviations calculated from at least two independent experiments. (*C*) Position of the cleavage sites on sense and antisense target RNAs. The target RNA sequences are as described in Figure 3B. Control 52-bp dsRNA (10 nM) or 21- and 22-nt RNA duplexes 1–7 (100 nM) were incubated with target RNA at 25°C for 2.5 h in *Drosophila* lysate. The stable 5′ cleavage products were resolved on the gel. The cleavage sites are indicated in Figure 5A. The region targeted by the 52-bp dsRNA or the sense (s) or antisense (as) strands are indicated by the black bars to the side of the gel. The cleavage sites are all located within the region of identity of the dsRNAs. For precise determination of the cleavage sites of the antisense strand, a lower percentage gel was used (data not shown).

downstream of the first nucleotide that is complementary to the 21- or 22-nt guide sequence (Fig. 4A,B). Displacing the sense strand of a 22-nt duplex by two nucleotides (cf. duplexes 1 and 3 in Fig. 5A) displaced the cleavage site of only the antisense target by two nucleotides. Displacing both sense and antisense strands by two nucleotides shifted both cleavage sites by two nucleotides (cf. duplexes 1 and 4). We predict that it will be possible to design a pair of 21- or 22-nt RNAs to cleave a target RNA at almost any given position.

The specificity of target RNA cleavage guided by 21- and 22-nt RNAs appears exquisite, as no cleavage sites are detected outside of the region of complementarity to the 21- and 22-nt RNAs (Fig. 5C). It should, however, be noted that the nucleotides present in the 3′ overhang of the 21- and 22-nt RNA duplex may contribute less to substrate recognition than the nucleotides near the cleavage site. This is based on the observation that the 3′-most nucleotide of the antisense strand of active duplexes 1 or 3 (Fig. 5A) is not complementary to the target. A detailed analysis of the specificity of RNAi can now be readily undertaken using synthetic 21- and 22-nt RNAs.

On the basis of the evidence that synthetic 21- and 22-nt RNAs with overhanging 3′ ends mediate RNA interference, we propose to name the ~21-nt RNAs short interfering RNAs, or siRNAs, and the respective RNA-protein complex a small interfering ribonucleoprotein particle, or siRNP.

3′ overhangs of ~20 nt on short dsRNAs inhibit RNAi

We have also analyzed dsRNAs with 17–20 nt overhanging 3′ ends that were less potent than blunt-ended dsRNAs (data not shown). The inhibitory effect of long 3′ ends was particularly pronounced for dsRNAs <100 bp. The effect was not caused by imperfect dsRNA formation, based on native gel analysis (data not shown). We tested if the inhibitory effect of long overhanging 3′ ends could be used as a tool to initiate dsRNA processing at only one of the two ends of a short RNA duplex.

We synthesized four combinations of the 52-bp model dsRNA, blunt-ended, 3′ extension only on the sense strand, the 3′ extension only on the antisense strand, and the double 3′ extension on both strands and mapped the target RNA cleavage sites and monitored ~21-nt formation after incubation in lysate (Fig. 6A–C). The first and predominant cleavage site of the sense target was lost when the 3′ end of the antisense strand of the duplex was extended, and the strong cleavage site of the antisense target was lost when the 3′ end of sense strand of the duplex was extended. Extending the 3′ ends on both strands rendered the 52-bp dsRNA virtually inactive. These observations correlate with the formation of ~21-nt fragments from blunt-ended dsRNAs or dsRNAs with only one 3′ extension and the absence of ~21-nt fragments when both 3′ ends of the duplex are extended (Fig. 6C). One explanation for the dsRNA inactivation by ~20-nt 3′ extensions could be the association of single-stranded RNA-binding proteins that could interfere with the association of one of the dsRNA-processing factors at this end. This is supported by the significantly longer persistence of the double 3′ extended dsRNA in the lysate (Fig. 6C). Together, these results are consistent with our model where only one of the strands of the siRNA duplex in the assembled siRNP is able to guide target RNA cleavage. The orientation of the strand that guides RNA cleavage is defined by the direction of the dsRNA processing reaction. A block at the 3′ end of the sense strand will only permit dsRNA processing from the opposing 3′ end of the antisense strand. This, in turn, generates siRNPs in which only the antisense strand of the siRNA duplex is able to guide sense target RNA cleavage. The same is true for the reciprocal situation. The less pronounced inhibitory effect of long 3′ extensions in the case of longer dsRNAs (≥500 bp, data not shown) suggests that long dsRNAs may also contain internal dsRNA-processing signals or may get processed cooperatively because of the association of multiple cleavage factors.

A model for dsRNA-directed mRNA cleavage

The new biochemical data update the model for how dsRNA targets mRNA for destruction (Fig. 7). Based on the 21–23-nt length of the processed RNA fragments, it had already been speculated that an RNase III–like activity may be involved in RNAi (Bass 2000). Double-stranded RNA is first processed to short RNA duplexes of predominantly 21 and 22 nt in length and with staggered 3′ ends similar to an RNase III–like reaction (Dunn 1982; Robertson 1982; Nicholson 1999). This hypothesis is further supported by the presence of 5′ phosphates and 3′ hydroxyls at the termini of the siRNAs (Fig. 4B) as observed in RNase III reaction products (Dunn 1982; Nicholson 1999).

Bacterial RNase III and the eukaryotic homologs Rnt1p in *Saccharomyces cerevisiae* and Pac1p in *Schizosaccharomyces pombe* have been shown to function in processing of ribosomal RNA as well as snRNA and snoRNAs (see, for example, Chanfreau et al. 2000). Little is known about the biochemistry of RNase III homologs from plants, animals, or human. Two families of RNase III enzymes have been identified predominantly by database-guided sequence analysis or cloning of cDNAs. The first RNase III family is represented by the 1327-amino-acid *D. melanogaster* protein *drosha* (accession no. AF116572). The carboxyl terminus is composed of two RNase III domains and one dsRNA-binding domain, and the amino terminus is of unknown function. Close homologs are also found in *C. elegans* (accession no. AF160248) and human (accession no. AF189011; Filippov et al. 2000; Wu et al. 2000). The *drosha*-like human RNase III was recently cloned and characterized (Wu et al. 2000). The gene is ubiquitously expressed in human tissues and cell lines, and the protein is localized in the nucleus and the nucleolus of the cell. Based on results inferred from antisense inhibition studies, a role of this protein for rRNA processing was suggested. The second class is represented by the *C. elegans* gene K12H4.8 (accession no. S44849) coding for an 1822-amino-acid pro-

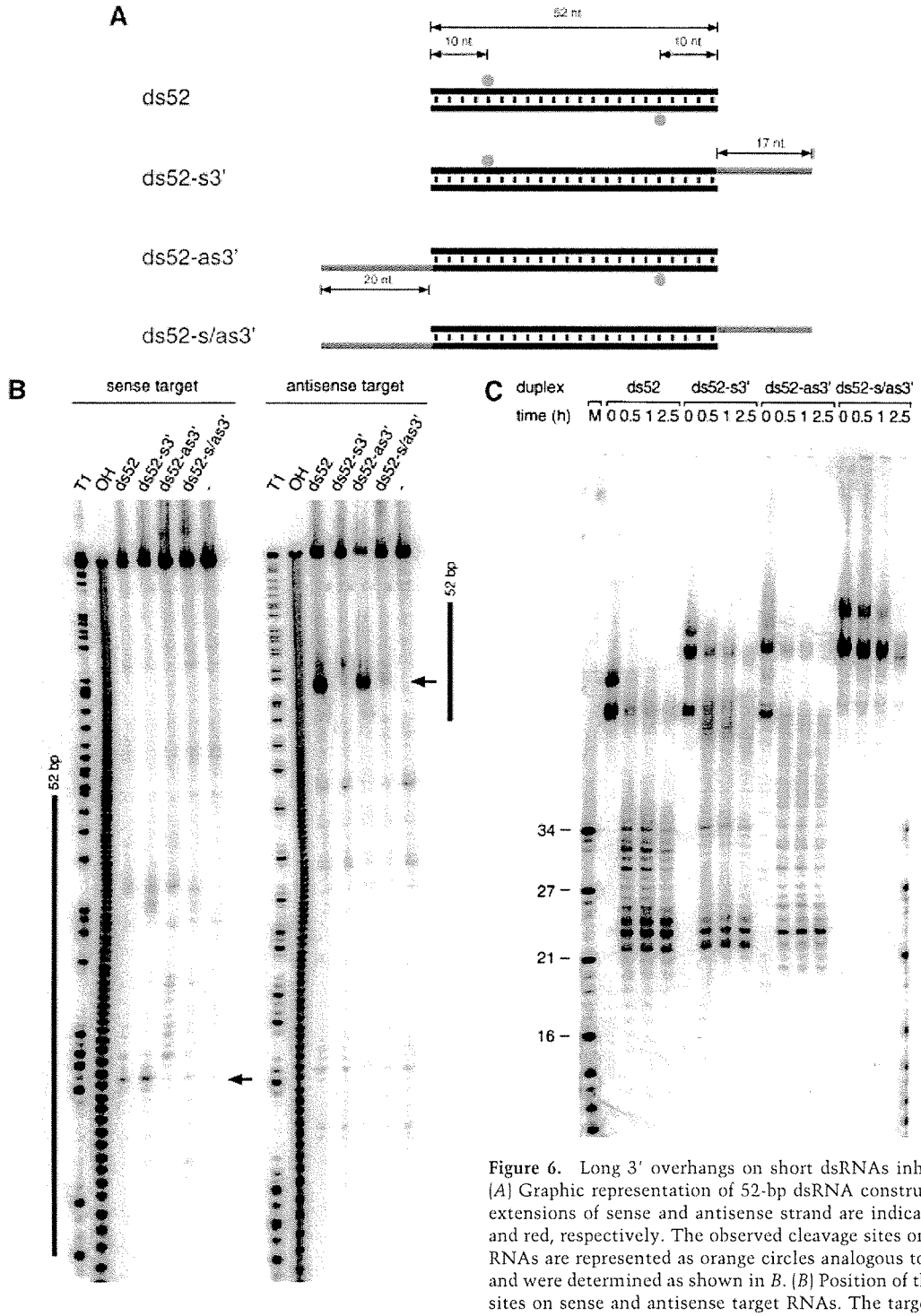


Figure 6. Long 3′ overhangs on short dsRNAs inhibit RNAi. (*A*) Graphic representation of 52-bp dsRNA constructs. The 3′ extensions of sense and antisense strand are indicated in blue and red, respectively. The observed cleavage sites on the target RNAs are represented as orange circles analogous to Figure 4A and were determined as shown in *B*. (*B*) Position of the cleavage sites on sense and antisense target RNAs. The target RNA sequences are as described in Figure 3B. DsRNA (10 nM) was incubated with target RNA at 25°C for 2.5 h in *Drosophila* lysate. The stable 5′ cleavage products were resolved on the gel. The major cleavage sites are indicated with a horizontal arrow and are also represented in *A*. The region targeted by the 52-bp dsRNA is represented as a black bar at both sides of the gel. (*C*) Processing of 52-bp dsRNAs with different 3′ extensions. Internally ^{32}P-labeled dsRNAs (5 nM) were incubated in *Drosophila* lysate, and reaction aliquots were analyzed at the indicated time points. An RNA size marker (M) has been loaded in the left lane, and the fragment sizes are indicated. Double bands at time zero are caused by incompletely denatured dsRNA.

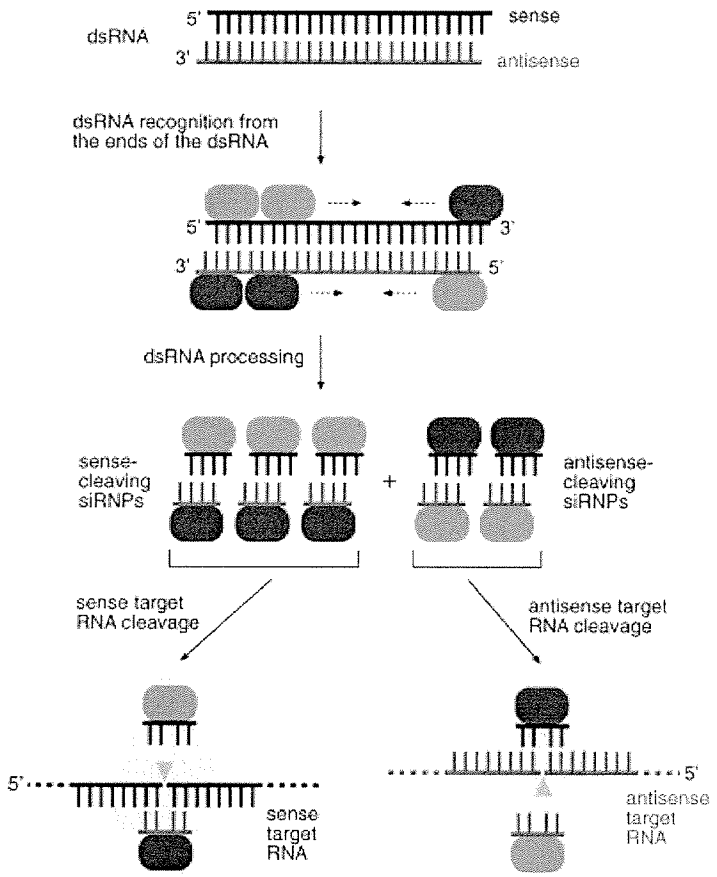


Figure 7. Proposed model for RNAi. RNAi is predicted to begin with processing of dsRNA (sense strand in black, antisense strand in red) to predominantly 21- and 22-nt short interfering RNAs (siRNAs). Short overhanging 3′ nucleotides, if present on the dsRNA, may be beneficial for processing of short dsRNAs. The dsRNA-processing proteins, which remain to be characterized, are represented as green and blue ovals and assemble on the dsRNA in asymmetric fashion. In our model, this is illustrated by binding of a hypothetical blue protein or protein domain with the siRNA strand in 3′ to 5′ direction while the hypothetical green protein or protein domain is always bound to the opposing siRNA strand. These proteins or a subset remain associated with the siRNA duplex and preserve its orientation, as determined by the direction of the dsRNA processing reaction. Only the siRNA sequence associated with the blue protein is able to guide target RNA cleavage. The endonuclease complex is referred to as small interfering ribonucleoprotein complex or siRNP. It is presumed here that the endonuclease that cleaves the dsRNA may also cleave the target RNA, probably by temporarily displacing the passive siRNA strand not used for target recognition. The target RNA is then cleaved in the center of the region recognized by the sequence-complementary guide siRNA. Because the cleavage site is displaced by 10–12 nt relative to the dsRNA processing site, a conformational rearrangement or a change in the composition of an siRNP must occur before target RNA cleavage.

tein. This protein has an amino-terminal RNA helicase motif, which is followed by two RNase III catalytic domains and a dsRNA-binding motif, similar to the *drosha* RNase III family. There are close homologs in *S. pombe* (accession no. Q09884), *Arabidopsis thaliana* (accession no. AF187317), *D. melanogaster* (accession no. AE003740), and human (accession no. AB028449) (Jacobsen et al. 1999; Filippov et al. 2000; Matsuda et al. 2000). It is tempting to speculate that the K12H4.8 RNase III/helicase is the likely candidate to be involved in RNAi.

Genetic screens in *C. elegans* identified *rde-1* and *rde-4* as essential for activation of RNAi, without an effect on transposon mobilization or cosuppression (Tabara et al. 1999; Dernburg et al. 2000; Grishok et al. 2000; Ketting and Plasterk 2000). This led to the hypothesis that these genes are important for dsRNA processing but are not involved in mRNA target degradation. The function of both genes is as yet unknown, the *rde-1* gene product is a member of a family of proteins similar to the rabbit protein eIF2C (Tabara et al. 1999), and the sequence of *rde-4* has not yet been described. Future biochemical characterization of these proteins should reveal their molecular function.

Processing of dsRNA to siRNA duplexes appears to start from the ends of both blunt-ended dsRNAs or dsRNAs with short (1–5 nt) 3′ overhangs and proceeds in ~21–23-nt steps. Long (~20 nt) 3′ staggered ends on short dsRNAs suppress RNAi, possibly through interaction with single-stranded RNA-binding proteins. The suppression of RNAi by single-stranded regions flanking short dsRNA and the reduced rate of siRNA formation from short 30-bp dsRNAs may explain why structured regions within mRNAs do not lead to activation of RNAi. In *C. elegans*, it was observed recently that injection of a 26-bp dsRNA could trigger RNAi of the *unc-22* gene; however, a >250-fold higher concentration of 26-bp dsRNA was necessary compared to an 81-bp dsRNA control (Parrish et al. 2000). It is conceivable that siRNA production from the 26-bp dsRNA was rate limiting and was compensated for by increasing the concentration of 26-bp dsRNA.

In our model, we presume that the dsRNA-processing proteins or a subset of them remain associated with the siRNA duplex after the processing reaction. The orientation of the siRNA duplex relative to these proteins determines which of the two complementary strands functions in guiding target RNA degradation. Chemically synthesized siRNA duplexes guide cleavage of sense as well as antisense target RNA, as they are able to associate with the protein components in either of the two possible orientations. A distinct role of the two strands of an siRNP is consistent with the recent obser-

vation in *C. elegans* that certain chemical modifications (e.g., 2′-aminouridine, 2′-deoxythymidine, or 5-iodouridine) incorporated into dsRNA are well tolerated at the sense, but not the cleavage-guiding antisense, strand (Parrish et al. 2000).

The finding that synthetic 21- and 22-nt siRNA duplexes can be used for efficient mRNA degradation demonstrates that the targeting step can be uncoupled from the dsRNA-processing step. This raises the prospects of using siRNA duplexes as new tools for sequence-specific regulation of gene expression in functional genomics as well as biomedical studies. The siRNAs may be effective in mammalian systems, where long dsRNAs can not be used because they activate the dsRNA-dependent protein kinase (PKR) response (Clemens 1997). As such, the siRNA duplexes may represent a new alternative to antisense or ribozyme therapeutics.

Materials and methods

In vitro RNAi

In vitro RNAi and lysate preparations were performed as described previously (Tuschl et al. 1999; Zamore et al. 2000) using a final concentration of 0.03 mg/mL creatine kinase in the RNAi reaction. It is critical to use freshly dissolved creatine kinase (Roche or Sigma) for optimal ATP regeneration. The RNAi translation assays (Fig. 1) were performed with dsRNA concentrations of 5 nM and an extended preincubation period at 25°C for 15 min before the addition of in vitro transcribed, capped, and polyadenylated *Pp*-luc and *Rr*-luc reporter mRNAs. The incubation was continued for 1 h, and the relative amount of *Pp*-luc and *Rr*-luc protein was analyzed using the dual luciferase assay (Promega) and a Monolight 3010C luminometer (PharMingen).

RNA synthesis

Standard procedures were used for in vitro transcription of RNA from PCR templates carrying T7 or SP6 promoter sequences (see, for example, Tuschl et al. 1998). Synthetic RNA was prepared using Expedite RNA phosphoramidites (Proligo). The 3′ adapter oligonucleotide was synthesized using dimethoxytrityl-1,4-benzenedimethanol-succinyl-aminopropyl-CPG, a generous gift from B. Sproat (Catholic University Leuven, Belgium). The oligoribonucleotides were deprotected in 3 mL of 32% ammonia/ethanol (3:1) at 55°C for 4 h (Expedite RNA) or at 55°C for 16 h (3′ and 5′ adapter DNA/RNA chimeric oligonucleotides) and then desilylated and gel purified as described previously (Tuschl et al. 1993). RNA transcripts for dsRNA preparation including long 3′ overhangs were generated from PCR templates that contained a T7 promoter in sense and an SP6 promoter in antisense direction. The transcription template for sense and antisense target RNA was PCR amplified with GCGTAATACGACTCACTATAGAACAATTGCTTTTACAG (underlined, T7 promoter) as 5′ primer, ATTTAGGTGACACTATAGGCATAAAGAATTGAAGA (underlined, SP6 promoter), as 3′ primer and the linearized *Pp*-luc plasmid (pGEM-luc sequence; Tuschl et al. 1999) as template; the T7-transcribed sense RNA was 177 nt long with the *Pp*-luc sequence between positions 113–273 relative to the start codon and followed by 17 nt of the complement of the SP6 promoter sequence at the 3′ end. Transcripts for blunt-ended dsRNA formation were prepared by transcription from two different PCR products that only contained a single promoter sequence.

DsRNA annealing was carried out using a phenol/chloroform extraction. Equimolar concentration of sense and antisense RNA (50 nM to 10 µM, depending on the length and amount available) in 0.3 M NaOAc (pH 6) were incubated at 90°C for 30 sec and then extracted at room temperature with an equal volume of phenol/chloroform and followed by a chloroform extraction to remove residual phenol. The resulting dsRNA was precipitated by addition of 2.5–3 volumes of ethanol. The pellet was dissolved in lysis buffer (100 mM KCl, 30 mM HEPES-KOH at pH 7.4, 2 mM $Mg(OAc)_2$), and the quality of the dsRNA was verified by standard agarose gel electrophoresis in 1× TAE-buffer. The 52-bp dsRNAs with the 17-nt and 20-nt 3′ overhangs (Fig. 6) were annealed by incubating at 95°C for 1 min and then were rapidly cooled to 70°C and followed by slow cooling to room temperature over a 3-h period (50 µL annealing reaction, 1 µM strand concentration, 300 mM NaCl, 10 mM Tris-HCl at pH 7.5). The dsRNAs were then phenol/chloroform extracted, ethanol precipitated, and dissolved in lysis buffer.

Transcription of internally ^{32}P-radiolabeled RNA used for dsRNA preparation (Figs. 2,4) was performed using 1 mM ATP, CTP, GTP; 0.1 or 0.2 mM UTP; and 0.2–0.3 µM [α-^{32}P]UTP (3000 Ci/mmol) or the respective ratio for radiolabeled nucleoside triphosphates other than UTP. Labeling of the cap of the target RNAs was performed as described previously (Zamore et al. 2000). The target RNAs were gel purified after cap labeling.

Cleavage site mapping

Standard RNAi reactions were performed by preincubating 10 nM dsRNA for 15 min followed by addition of 10 nM cap-labeled target RNA. The reaction was stopped after a further 2-h (Fig. 3A) or 2.5-h incubation (Figs. 5C,6B) by proteinase K treatment (Tuschl et al. 1999). The samples were then analyzed on 8% or 10% sequencing gels. The 21- and 22-nt synthetic RNA duplexes were used at 100 nM final concentration (Fig. 5B,C).

Cloning of ~21-nt RNAs

The 21-nt RNAs were produced by incubation of radiolabeled dsRNA in *Drosophila* lysate in absence of target RNA (200 µL reaction, 1 h incubation, 50 nM dsP111, or 100 nM dsP52 or dsP39). The reaction mixture was subsequently treated with proteinase K (Tuschl et al. 1999), and the dsRNA-processing products were separated on a denaturing 15% polyacrylamide gel. A band, including a size range of at least 18–24 nt, was excised and then eluted into 0.3 M NaCl overnight at 4°C in siliconized tubes. The RNA was recovered by ethanol precipitation and then dephosphorylated (30 µL reaction, 50°C, 30 min, 10 U alkaline phosphatase; Roche). The reaction was stopped by phenol/chloroform extraction, and the RNA was ethanol precipitated. The 3′ adapter oligonucleotide (pUUUaaccg catccttctcx: uppercase, RNA; lowercase, DNA; p, phosphate; x, 4-hydroxymethylbenzyl) was then ligated to the dephosphorylated ~21-nt RNA (20 µL reaction, 37°C, 30 min, 5 µM 3′ adapter, 50 mM Tris-HCl at pH 7.6, 10 mM $MgCl_2$, 0.2 mM ATP, 0.1 mg/mL acetylated BSA, 15% DMSO, 25 U T4 RNA ligase; Amersham-Pharmacia) (Pan and Uhlenbeck 1992). The ligation reaction was stopped by the addition of an equal volume of 8 M urea/50 mM EDTA stopmix and directly loaded on a 15% gel. Ligation yields were >50%. The ligation product was recovered from the gel and 5′ phosphorylated (20 µL reaction, 37°C, 30 min, 2 mM ATP, 5 U T4 polynucleotide kinase; NEB). The phosphorylation reaction was stopped by phenol/chloroform extraction, and RNA was recovered by ethanol precipita-

tion. Next, the 5′ adapter (tactaatacgactcactAAA: uppercase, RNA; lowercase, DNA) was ligated to the phosphorylated ligation product as described above. The new ligation product was gel purified and eluted from the gel slice in the presence of reverse transcription primer (GACTAGCTG**GAATTC**AAG GATGCGGTTAAA: bold, *Eco*RI site), used as carrier. Reverse transcription (15 µL reaction, 42°C, 30 min, 150 U Superscript II reverse transcriptase; Life Technologies) was followed by PCR using a 5′ primer CAGCCAACG**GAATTC**ATACGACTCAC TAAA (bold, *Eco*RI site) and the 3′ RT primer. The PCR product was purified by phenol/chloroform extraction and ethanol precipitated. The PCR product was then digested with *Eco*RI (NEB) and concatamerized using T4 DNA ligase (high concentration; NEB). Concatamers of a size range of 200–800 bp were separated on a low-melt agarose gel, recovered from the gel by a standard melting and phenol extraction procedure, and ethanol precipitated. The unpaired ends were filled in by incubation with Taq polymerase under standard conditions at 72°C for 15 min, and the DNA product was directly ligated into the pCR2.1-TOPO vector using the TOPO TA cloning kit (Invitrogen). Colonies were screened using PCR and M13–20 and M13 Reverse sequencing primers. PCR products were directly submitted for custom sequencing (Sequence Laboratories Göttingen). On average, four to five ~21-mer sequences were obtained per clone.

2D-TLC analysis

Nuclease P1 digestion of radiolabeled, gel-purified siRNAs and 2D-TLC was carried out as described (Zamore et al. 2000). Nuclease T2 digestion was performed in 10 µL reactions at 50°C for 3 h in 10 mM ammonium acetate (pH 4.5) using 2 µg/µL carrier tRNA and 30 U ribonuclease T2 (Life Technologies). The migration of nonradioactive standards was determined by UV shadowing. The identity of nucleoside-3′,5′-diphosphates was confirmed by comigration of the T2 digestion products with standards prepared by 5′-^{32}P-phosphorylation of commercial nucleoside 3′-monophosphates using [γ-^{32}P]ATP and T4 polynucleotide kinase (data not shown).

Acknowledgments

We acknowledge Heike Taubner for assistance with fly work; Uschi Kutzke for chemical RNA synthesis; Gordon Dowe for some of the sequencing; and H. Jäckle, R. Lührmann, and F. Eckstein for support. We thank N.J. Watkins, T. Achsel, J. Ludwig, P.D. Zamore, D.P. Bartel, and P.A. Sharp for advice and comments on the manuscript. I would also like to thank P.D.Z and P.A.S. for the suggestion of naming the short interfering 21- and 22-nt RNAs siRNAs. This work was supported by BMBF Biofuture grant number 0311856.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Anandalakshmi, R., Marathe, R., Ge, X., Herr, Jr., J.M., Mau, C., Mallory, A., Pruss, G., Bowman, L., and Vance, V.B. 2000. A calmodulin-related protein that suppresses posttranscriptional gene silencing in plants. *Science* **290:** 142–144.

Bass, B.L. 2000. Double-stranded RNA as a template for gene silencing. *Cell* **101:** 235–238.

Bosher, J.M. and Labouesse, M. 2000. RNA interference: Genetic wand and genetic watchdog. *Nat. Cell Biol.* **2:** E31–E36.

Caplen, N.J., Fleenor, J., Fire, A., and Morgan, R.A. 2000. dsRNA-mediated gene silencing in cultured *Drosophila* cells: A tissue culture model for the analysis of RNA interference. *Gene* **252:** 95–105.

Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. 2000. Gene silencing in worms and fungi. *Nature* **404:** 245.

Chanfreau, G., Buckle, M., and Jacquier, A. 2000. Recognition of a conserved class of RNA tetraloops by *Saccharomyces cerevisiae* RNase III. *Proc. Natl. Acad. Sci.* **97:** 3142–3147.

Clemens, M.J. 1997. PKR—A protein kinase regulated by double-stranded RNA. *Int. J. Biochem. Cell Biol.* **29:** 945–949.

Cogoni, C. and Macino, G. 1999. Homology-dependent gene silencing in plants and fungi: A number of variations on the same theme. *Curr. Opin. Microbiol.* **2:** 657–662.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. 2000. An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101:** 543–553.

Dernburg, A.F., Zalevsky, J., Colaiacovo, M.P., and Villeneuve, A.M. 2000. Transgene-mediated cosuppression in the *C. elegans* germ line. *Genes & Dev.* **14:** 1578–1583.

Dunn, J.J. 1982. Ribonuclease III. In *The enzymes*, Vol. 15, part B. (ed. P.D. Boyer), pp. 485–499. Academic Press, New York.

Filippov, V., Solovyev, V., Filippova, M., and Gill, S.S. 2000. A novel type of RNase III family proteins in eukaryotes. *Gene* **245:** 213–221.

Fire, A. 1999. RNA-triggered gene silencing. *Trends Genet.* **15:** 358–363.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391:** 806–811.

Grishok, A., Tabara, H., and Mello, C.C. 2000. Genetic requirements for inheritance of RNAi in *C. elegans*. *Science* **287:** 2494–2497.

Hamilton, A.J. and Baulcombe, D.C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286:** 950–952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. 2000. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* **404:** 293–296.

Jacobsen, S.E., Running, M.P., and Meyerowitz, M.E. 1999. Disruption of an RNA helicase/RNase III gene in *Arabidopsis* causes unregulated cell division in floral meristems. *Development* **126:** 5231–5243.

Jensen, S., Gassama, M.P., and Heidmann, T. 1999. Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* **21:** 209–212.

Kennerdell, J.R. and Carthew, R.W. 1998. Use of dsRNA-mediated genetic interference to demonstrate that *frizzled* and *frizzled 2* act in the wingless pathway. *Cell* **95:** 1017–1026.

Ketting, R.F. and Plasterk, R.H. 2000. A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* **404:** 296–298.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. 1999. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* **99:** 133–141.

Lucy, A.P., Guo, H.S., Li, W.X., and Ding, S.W. 2000. Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. *EMBO J.* **19:** 1672–1680.

Malinsky, S., Bucheton, A., and Busseau, I. 2000. New insights

on homology-dependent silencing of I factor activity by transgenes containing ORF1 in *Drosophila melanogaster*. *Genetics* **156:** 1147–1155.

Matsuda, S., Ichigotani, Y., Okuda, T., Irimura, T., Nakatsugawa, S., and Hamaguchi, M. 2000. Molecular cloning and characterization of a novel human gene (HERNA) which encodes a putative RNA-helicase. *Biochim. Biophys. Acta* **1490:** 163–169.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., et al. 2000. *Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* **101:** 533–542.

Ngo, H., Tschudi, C., Gull, K., and Ullu, E. 1998. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc. Natl. Acad. Sci.* **95:** 14687–14692.

Nicholson, A.W. 1999. Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol. Rev.* **23:** 371–390.

Oelgeschlager, M., Larrain, J., Geissert, D., and De Robertis, E.M. 2000. The evolutionarily conserved BMP-binding protein Twisted gastrulation promotes BMP signalling. *Nature* **405:** 757–763.

Pan, T. and Uhlenbeck, O.C. 1992. In vitro selection of RNAs that undergo autolytic cleavage with Pb2+. *Biochemistry* **31:** 3887–3895.

Parrish, S., Fleenor, J., Xu, S., Mello, C., and Fire, A. 2000. Functional anatomy of a dsRNA trigger: Differential requirement for the two trigger strands in RNA Interference. *Mol. Cell* **6:** 1077–1087.

Plasterk, R.H. and Ketting, R.F. 2000. The silence of the genes. *Curr. Opin. Genet. Dev.* **10:** 562–567.

Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. 1999. Gene silencing without DNA: RNA-mediated cross-protection between viruses. *Plant Cell* **11:** 1207–1216.

Robertson, H.D. 1982. *Escherichia coli* ribonuclease III cleavage sites. *Cell* **30:** 669–672.

———. 1990. *Escherichia coli* ribonuclease III. *Methods Enzymol.* **181:** 189–202.

Romaniuk, E., McLaughlin, L.W., Neilson, T., and Romaniuk, P.J. 1982. The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur. J. Biochem.* **125:** 639–643.

Sharp, P.A. 1999. RNAi and double-strand RNA. *Genes & Dev.* **13:** 139–141.

Sijen, T. and Kooter, J.M. 2000. Post-transcriptional gene-silencing: RNAs on the attack or on the defense? *Bioessays* **22:** 520–531.

Smardon, A., Spoerke, J., Stacey, S., Klein, M., Mackin, N., and Maine, E. 2000. EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* **10:** 169–178.

Svoboda, P., Stein, P., Hayashi, H., and Schultz, R.M. 2000. Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference. *Development* **127:** 4147–4156.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99:** 123–132.

Tuschl, T., Ng, M.M., Pieken, W., Benseler, F., and Eckstein, F. 1993. Importance of exocyclic base functional groups of central core guanosines for hammerhead ribozyme activity. *Biochemistry* **32:** 11658–11668.

Tuschl, T., Sharp, P.A., and Bartel, D.P. 1998. Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. *EMBO J.* **17:** 2637–2650.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes & Dev.* **13:** 3191–3197.

Voinnet, O., Lederer, C., and Baulcombe, D.C. 2000. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell* **103:** 157–167.

Wianny, F. and Zernicka-Goetz, M. 2000. Specific interference with gene function by double-stranded RNA in early mouse development. *Nat. Cell Biol.* **2:** 70–75.

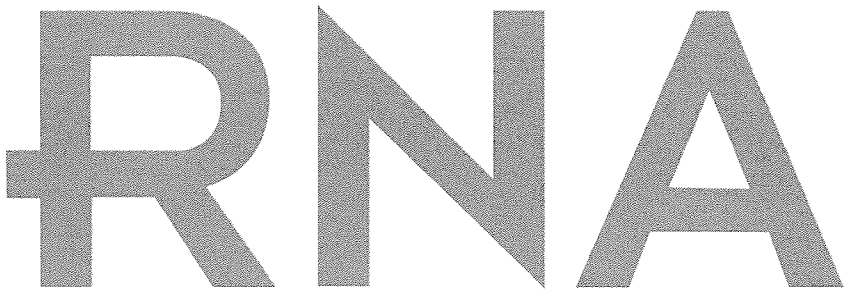
Wu, H., Xu, H., Miraglia, L.J., and Crooke, S.T. 2000. Human RNase III is a 160 kDa protein involved in preribosomal RNA processing. *J. Biol. Chem.* **17:** 17.

Yang, D., Lu, H., and Erickson, J.W. 2000. Evidence that processed small dsRNAs may mediate sequence-specific mRNA degradation during RNAi in *Drosophila* embryos. *Curr. Biol.* **10:** 1191–1200.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. 2000. RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101:** 25–33.

Zhang, K. and Nicholson, A.W. 1997. Regulation of ribonuclease III processing by double-helical sequence antideterminants. *Proc. Natl. Acad. Sci.* **94:** 13437–13441.

Downloaded from www.rnajournal.org on March 27, 2007



Activation of the protein kinase PKR by short double-stranded RNAs with single-stranded tails

XIAOFENG ZHENG and PHILIP C. BEVILACQUA

RNA 2004 10: 1934-1945
Access the most recent version at doi:10.1261/rna.7150804

References This article cites 45 articles, 24 of which can be accessed free at:
http://www.rnajournal.org/cgi/content/full/10/12/1934#References

Article cited in:
http://www.rnajournal.org/cgi/content/full/10/12/1934#otherarticles

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

Notes

To subscribe to *RNA* go to:
http://www.rnajournal.org/subscriptions/

© 2004 RNA Society

MOLECULAR AND CELLULAR BIOLOGY, Dec. 1991, p. 5848–5859
0270-7306/91/125848-12$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 11, No. 12

The Encephalomyocarditis Virus Internal Ribosome Entry Site Allows Efficient Coexpression of Two Genes from a Recombinant Provirus in Cultured Cells and in Embryos

INGRID R. GHATTAS,[1] JOSHUA R. SANES,[2] AND JOHN E. MAJORS[1]*

Departments of Biochemistry and Molecular Biophysics[1] and of Anatomy and Neurobiology,[2] Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110

Received 16 June 1991/Accepted 3 September 1991

Rous sarcoma virus-based retroviral vectors were constructed to compare three different approaches for coexpressing two genes in individual infected cells. All vectors expressed the upstream gene (*lacZ*) from the Rous sarcoma virus long terminal repeat, while the downstream gene (the chloramphenicol acetyltransferase gene [*cat*] or v-*src*) was expressed in one of three ways: from a subgenomic mRNA generated by regulated splicing, from a strong internal promoter, or from the encephalomyocarditis virus internal ribosome entry site (IRES). Both biochemical and immunohistochemical assays of cultured cells showed that the encephalomyocarditis virus IRES provided the most efficient means for coexpressing two genes from a single provirus. Most importantly, most cells infected by a LacZ-IRES-CAT virus expressed both LacZ and CAT, whereas most cells infected by internal promoter or regulated splicing vectors expressed either LacZ or CAT but not both. In addition, viral titers were highest with IRES vectors. Presumably, use of the IRES avoids transcriptional controls and RNA processing steps that differentially affect expression of multiple genes from internal promoter and regulated splicing vectors. Finally, we injected a LacZ–IRES–v-Src virus into chicken embryos and then identified the progeny of infected cells with a histochemical stain for LacZ. LacZ-positive cells in both skin and mesenchyme displayed morphological abnormalities attributable to expression of v-*src*. Thus, IRES vectors can be used to coexpress a reporter gene and a bioactive gene in vivo.

When a retrovirus infects a cell, the viral genome is stably integrated into the host chromosome, efficiently expressed, and faithfully passed to the infected cell's progeny. For these reasons, recombinant retroviral vectors have often been used to express exogenous genes in vertebrate cells (reviewed in references 28, 34, and 52). In some of these cases, it is advantageous to express two exogenous genes from a single proviral genome. In strategies being developed for gene therapy, for example, the retrovirus often contains not only the gene of interest but also a selectable marker. The marker is used to facilitate the isolation of infected cells, which are then used as a source of the potentially therapeutic gene product (reviewed in reference 12).

In another set of studies, we and others have used vectors encoding the histochemical marker β-galactosidase, the product of the *Escherichia coli lacZ* gene, as lineage tracers in vivo. A single cell is infected by a retrovirus, the proviral genome is inherited by the cell's progeny, and the clonal relatives are identified with the histochemical stain for LacZ (11, 13, 16, 18, 45; reviewed in reference 17). To extend this work, we wished to construct an efficient double-expression vector to transfer both *lacZ* and a second gene to single cells in vivo. If *lacZ* and a second bioactive gene were reliably coexpressed at high levels, we could use LacZ histochemistry to identify small clones of transgenic cells in a wild-type environment and then seek cell autonomous effects of the second gene by analyzing the number, distribution, and morphology of the labeled cells.

In applications such as these, it is essential that the provirus express both genes within the same individual cells. Retroviruses have been successful in evolving strategies for coexpressing their own genes. These strategies include synthesis and subsequent processing of fusion proteins, ribosome frameshifting, and regulated splicing to generate subgenomic messages. Unfortunately, achieving balanced expression of multiple exogenous genes from engineered retrovirus vectors has been more problematic. Two approaches have been used previously. The first involves the generation of separate mRNAs by regulated splicing of a single primary transcript expressed from the upstream long terminal repeat (LTR). This strategy mimics that used by retroviruses to generate the *env* gene product (54). With this approach, expression of one gene is always at the expense of the other, and the ratio of spliced to unspliced mRNA is highly dependent on the context (1, 2, 48, 49). The second approach involves expression of the upstream gene from the retrovirus promoter in the LTR and expression of the downstream gene from an internal promoter. This approach has been used most often in vectors designed for gene therapy (reviewed in references 28 and 34) but is compromised by competitive interference between promoters (4, 8, 20, 39). Thus, individual isolates may express one gene or the other, rather than both.

The recent demonstration that the 5′ nontranslated region of encephalomyocarditis virus (EMCV) and other picornaviruses allows translation initiation from the internal ribosome entry site (IRES) (22–24) has suggested a third strategy for expressing multiple genes from a single proviral genome: the viral LTR can be used to express a single polycistronic transcript from which several gene products are translated. In this way, transcriptional controls and RNA processing steps that differentially affect expression of multiple genes can be avoided. In this study, we have used three pairs of genes to compare this approach with the double-promoter and regulated-splicing strategies. Biochemical and immuno-

* Corresponding author.

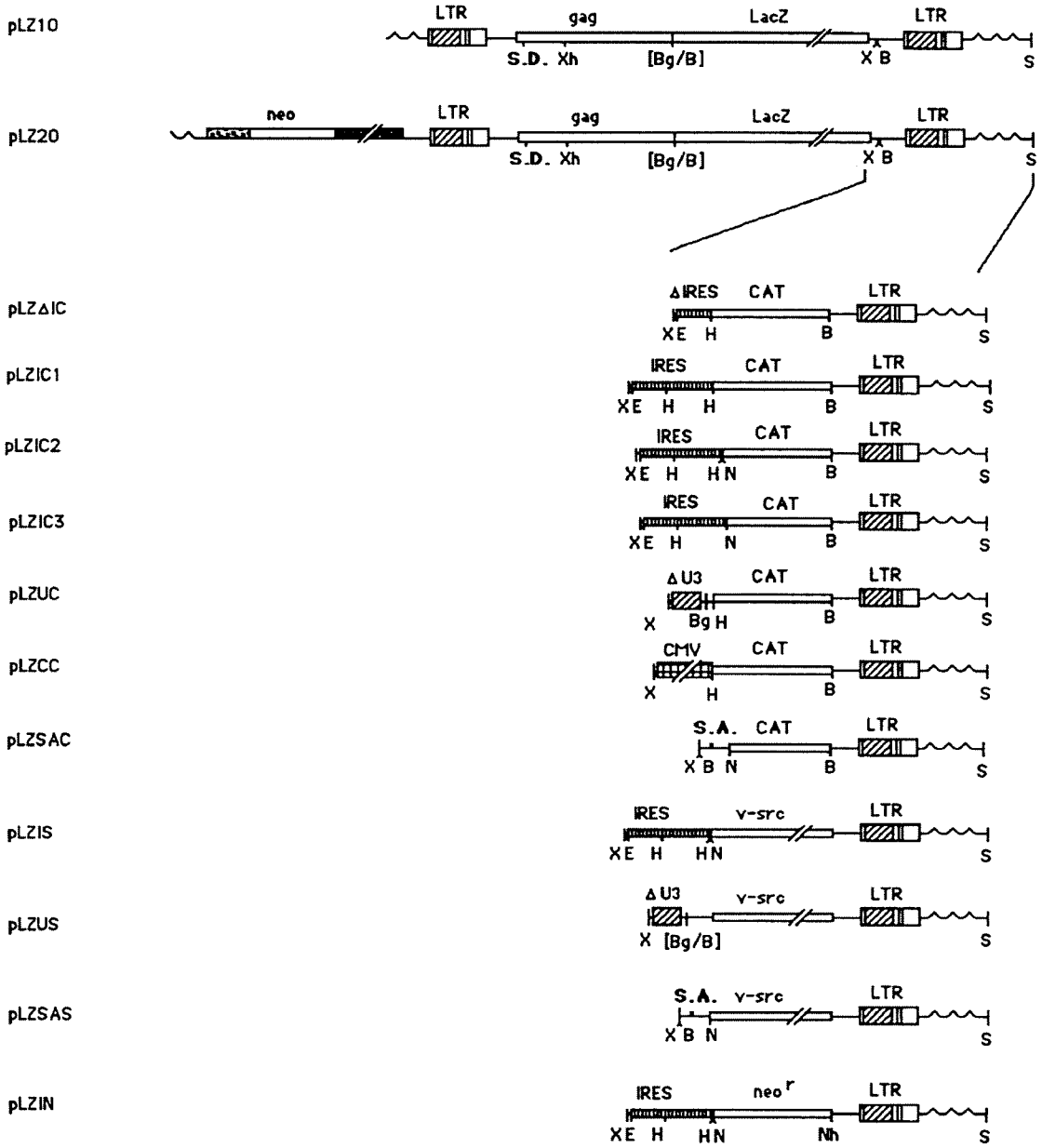


FIG. 1. Structures of retrovirus vectors used in this study. Notation: straight lines, RSV sequences; wavy lines, bacterial sequences; diagonal lines, truncated U3 region of RSV; squares, CMV immediate-early promoter; vertical lines, sequences of the EMCV 5′ nontranslated region; spotted and dark stippled boxes, SV40 promoter and SV40 poly(A) signal, respectively; S.A., splice acceptor site; S.D., splice donor site; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; N, *Nco*I; Nh, *Nhe*I; S, *Sal*I; X, *Xba*I; Xh, *Xho*I. Brackets indicate sites destroyed by ligation. The diagrams are roughly to scale.

histochemical assays indicate that the EMCV IRES is more effective than either the double-promoter or the regulated-splicing strategy at expressing two exogenous genes from a single proviral genome. Using an EMCV IRES vector that encodes LacZ and v-Src, we then infected chicken embryos in ovo. We show that LacZ-positive cells in the skin and meninges consistently express morphological abnormalities that are attributable to v-*src*.

MATERIALS AND METHODS

Plasmid constructions. DNA manipulations were done by standard methods (44). The main features of the plasmids described below are illustrated in Fig. 1.

All vector plasmids are derivatives of pLZ20, which is similar to the previously described pLZ10 (13) except for the addition within the plasmid of a Tn*5* neomycin resistance gene (*neo*) flanked by a simian virus 40 (SV40) promoter and

poly(A) signal. Both pLZ10 and pLZ20 are derived from the SRA-2 molecular clone of the Rous sarcoma virus (RSV) genome (50) and produce LacZ as a freely soluble Gag-LacZ fusion protein. For each new vector, the *Xba*I-*Sal*I fragment of pLZ20 was replaced by a different fragment of interest. Plasmid pLZΔIC contains a truncated EMCV IRES extending from the *Eco*RI site at nucleotide (nt) 2334 to the *Hin*dIII site at nt 2566 of pE5LVP0 (40); nt 2334 of the plasmid is equivalent to nt 260 of the EMCV 5′ nontranslated region. This fragment was linked to a 773-bp *Hin*dIII-*Bam*HI fragment of pSV2CAT which contains the bacterial chloramphenicol acetyltransferase (*cat*) gene (15). The IRES fragment in pLZIC1 extends from nt 2334 to the EMCV initiation codon at nt 2915, which was converted to a *Hin*dIII site (conversion from AAUAUGGCC to AAGCUUGCC extending from nt 2912 to 2920 of pE5LVP0) by site-directed mutagenesis (36). The oligonucleotide ACACGAATGATA AGCTTGCCACAACCA used for mutagenesis was obtained from the protein chemistry laboratory at Washington University. In pLZIC2, pLZIC3, pLZIS, and pLZIN, the EMCV sequence extends to an *Nco*I site 9 bp downstream of the EMCV initiation codon, which was altered in the cases of pLZIC2, pLZIS, and pLZIN. The *cat* gene in pLZIC2 and pLZIC3 has a 29-bp sequence of the 5′ nontranslated *cat* region deleted and the initiation codon converted to an *Nco*I site. The *Xba*I-*Bgl*II fragment in pLZUC and pLZUS includes a truncated RSV U3 region, extending from nt -220 (an artificial *Bam*HI site) to -10 (a *Taq*I site), joined to a short linker sequence ATACCGTCCCAGATCT. To generate pLZUC, this fragment was linked to the *Hin*dIII end of the *cat* fragment by using a 56-bp *Sfi*I-*Hin*dIII filler fragment derived from the SV40 T-antigen leader sequence that was modified at the *Sfi*I end by addition of a *Bgl*II linker. pLZCC was derived from pLZIC1 by substitution of the *Xba*I-*Hin*dIII EMCV fragment with a 780-bp *Xba*I-*Hin*dIII fragment that harbors the enhancer/promoter region of a cytomegalovirus (CMV) immediate-early gene (3). The v-*src* splice acceptor site used in pLZSAC is derived from the region upstream of SRA v-*src* and was transferred as a *Bam*HI-*Nco*I fragment from a previously described construction (21). The v-*src* gene in pLZIS, pLZUS, and pLZSAS was derived from this same construction and was transferred as a *Bam*HI-*Eco*RI fragment into pLZUC and pLZSAC to make pLZUS and pLZSAS and as an *Nco*I-*Eco*RI fragment to make pLZIS. Finally, the *Nco*I-*Nhe*I fragment of pLZIN, which includes the *neo* gene, was generated by the polymerase chain reaction, using primers which converted the *neo* AUG to an *Nco*I site and introduced an *Nhe*I site immediately downstream of the translation termination site. Joining to the 3′ LTR was effected by introduction of an *Nhe*I linker at the *Bam*HI site that separates the CAT sequences in pLZIC1 from the downstream viral sequences. The helper plasmid employed in these experiments used the CMV promoter fragment described above to promote expression of SRA viral sequences that extended from a *Sac*I site 120 bp 5′ to the *gag* AUG (6) to an *Rsa*I site 6 bp 3′ to the end of *env* (5). A polyadenylation signal was provided from the herpes simplex virus *tk* gene (32).

Transfections and virus recovery. These experiments used QT6 cells, a chemically transformed quail fibroblast cell line (37). Cells were grown at 37°C and 5% CO_2 in Earle's 199 medium supplemented with 5% tryptose phosphate buffer, 5% fetal bovine serum, 1% dimethyl sulfoxide, penicillin, and streptomycin. Cells were plated at 2×10^6 cells per 60-mm dish 24 h prior to transfection and were fed fresh medium 1 h before transfection. The cells were transfected by the calcium phosphate procedure (53) with one of the recombinant plasmids and the helper plasmid, which provided the viral genes necessary to produce viral particles. The calcium phosphate coprecipitates were formed by using 6 μg each of the recombinant and helper plasmids. Seven hours after the precipitate had been added, the cells were treated with dimethyl sulfoxide as described by Lopata et al. (31). The cells were then incubated for 40 h before the virus-containing media were harvested and cell extracts were made.

The harvested media were passed through a 0.45-μm-pore-size filter to remove intact cells, and Polybrene was added to a final concentration of 8 μg/ml to enhance the efficiency of infection. The virus-containing media were then added to fresh QT6 cells, which were grown for at least 48 h before expression of virus-encoded genes was assessed as described below.

Enzyme assays. Extracts of transfected cells were prepared by three cycles of freezing and thawing in 100 μl of 0.25 M Tris-HCl (pH 8.0), followed by centrifugation at 4°C for 5 min to remove cell debris. LacZ activity was measured by the method of Miller (35), with the amounts of extract and time of incubation adjusted so that the A_{420} was between 0.2 and 1. CAT activity was measured by the method of Gorman et al. (15). Briefly, extract of transfected cells was added to a reaction mixture containing 37.5 μl of Tris-HCl (pH 7.5), 1 μl of 50-mg/ml chloramphenicol, and 0.1 μCi of [^{14}C]chloramphenicol (60 mCi/mmol; New England Nuclear Corp.). The mixture was adjusted to a final volume of 135 μl and preincubated for 5 min at 37°C. The reaction was then started by adding 15 μl of 50 mM acetyl coenzyme A. The amount of extract was adjusted so that the percent conversion of the nonacetylated form to the acetylated form of chloramphenicol was between 10 and 50%. For CAT assays on infected cells, unlabeled chloramphenicol was omitted.

Histochemistry. For histochemical staining of cells expressing the LacZ gene product, cells were fixed in 2% formaldehyde in phosphate-buffered saline (PBS) for 10 min. The cells were then stained for LacZ at room temperature, overnight, in a mixture containing 1 mg of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) per ml, 4 mM potassium ferricyanide, 4 mM potassium ferrocyanide, and 2 mM $MgCl_2$ in PBS (45). For immunohistochemical detection of LacZ and CAT, infected cells were seeded on coverslips 24 h prior to staining. The cells were fixed in 1% formaldehyde in PBS for 30 min and then incubated with antibodies. Primary antibodies were mouse monoclonal anti-β-galactosidase, prepared in our laboratories, rabbit anti-CAT (5 Prime → 3 Prime Inc., West Chester, Pa.), and rabbit antiserum to bacterially produced v-Src, generously provided by P. Maness, University of North Carolina (33). Secondary antibodies were (i) fluorescein-conjugated goat anti-mouse immunoglobulin G (IgG) plus IgM and (ii) rhodamine-conjugated goat anti-rabbit IgG (Boehringer Mannheim, Indianapolis, Ind.). For LacZ and CAT, each incubation was for 30 min at room temperature in PBS containing 5% goat serum and 0.01% Triton X-100. The v-*src* gene product was detected by incubating the cells overnight at 4°C with primary antibody, followed by a 30-min incubation with biotin-conjugated anti-rabbit IgG (Sigma, St. Louis, Mo.) and a 30-min incubation with Texas red-conjugated streptavidin (Bethesda Research Laboratories). Finally, coverslips were washed with PBS and mounted in 80% glycerol–*p*-phenylenediamine.

Infection of embryos. Methods for inoculation of embryos

and processing of tissue have been described previously (13, 16, 29). Briefly, fertilized White Leghorn eggs were obtained from SPAFAS (Roanoke, Ill.) and incubated at 37°C. Embryos were exposed through a window in the shell and staged by the criteria of Hamburger and Hamilton (19). Virus was concentrated from the medium of virus-producing QT6 cells by centrifugation, mixed with fast green and Polybrene, and drawn into a 10- to 30-μm-tip glass capillary microelectrode. A total of 0.5 to 3 μl of this mixture was pressure injected over the skin, subdermally, and into the lumen of the neural tube. No attempt was made to limit the injections to particular regions of the embryo. The shell was then closed with tape, and the egg was returned to the incubator.

At appropriate times thereafter, embryos were recovered, eviscerated, fixed for 1 h with 2% formaldehyde plus 0.4% gluteraldehyde in PBS, washed thoroughly in PBS, incubated overnight in X-Gal solution (see above), rinsed in PBS, and refixed in 2% formaldehyde–2% gluteraldehyde. The embryos were then examined under a dissecting microscope at ×20 to ×40. Small pieces of tissue bearing LacZ-positive cells were cut out and either mounted whole in 90% glycerol–10% PBS or dehydrated in ethanol, embedded in Araldite, and sectioned at 5 μm.

RESULTS

In this study, we compared three different strategies for coexpressing *lacZ* and a downstream gene from RSV-based retrovirus vectors. The upstream gene was always fused in frame with the *gag* polyprotein and expressed from the LTR of the SRA-2 clone of RSV (50). The downstream gene was expressed in one of three ways: from a spliced subgenomic message, from an internal promoter, or from the EMCV IRES. To generate spliced mRNA from primary transcripts, we used the splice acceptor site of the RSV v-*src* gene (51). In these vectors, the RSV LTR promotes expression of primary transcripts, which are either used intact for packaging and for translation of the upstream gene or spliced to form a subgenomic RNA from which the downstream gene is translated. In cells infected with the wild-type RSV, about half of the primary transcripts are spliced to generate *env* and v-*src* subgenomic mRNAs, although the precise ratio of spliced to unspliced RNA varies among strains (49). For an internal promoter, we used a fragment of the U3 region of the RSV LTR (50). This fragment includes promoter and enhancer sequences but excludes both the inverted repeat sequence at the 5′ end, which is required for integration, and the polyadenylation signal at the 3′ end. In one vector, we also used the immediate-early promoter of CMV as an internal promoter. Finally, a fragment of EMCV that contains an IRES was used to promote internal ribosome binding. *lacZ* was paired with three genes for these tests: *cat*, because a simple and sensitive test is available for its products; v-*src*, because we expected it to alter cellular morphology in ways that would be detectable in vivo; and *neo*, because it is a selectable marker that can be used in studies of transcriptional activation.

Structure of the EMCV IRES fragment. Translation of EMCV mRNA occurs in a cap-independent fashion: ribosomes bind internally at the initiating AUG without scanning the 5′ nontranslated region of the transcript. Sequences between nt 403 and 811 of the EMCV 5′ nontranslated region are required for efficient translation. They include a stem-loop structure which interacts with a 57-kDa cellular protein and a stretch of pyrimidine-rich sequences near the initiation codon, both of which are essential for IRES function (24). Synthesis of the EMCV polyprotein initiates at the 11th AUG codon, which is located in a perfect Kozak consensus sequence (27).

In earlier studies which used the IRES, the *cat* gene was expressed either from the EMCV initiation codon as a fusion protein (22) or from an engineered *Nco*I site placed at the EMCV initiation codon (7). We wanted to construct a vector in which foreign genes could be expressed either from their own initiation codons or from one in the vector. To do so, we took note that the 12th EMCV AUG, located 9 bp downstream of the EMCV initiation codon (the 11th AUG), is part of an *Nco*I site and is surrounded by a Kozak consensus sequence. To assess whether we could use that site as an initiation codon and to provide a site for insertion of genes with their own initiation codons, we converted the natural translation start site of the EMCV to a *Hin*dIII site. Thus, the sequence 5′-AAU AUG GCC ACA ACC AUG GAA-3′ was mutated to 5′-AAG CUU GCC ACA ACC AUG GAA-3′. The efficiency of translation from the 12th AUG was tested by placing the *lacZ* gene at that site in the wild-type context and in the altered context and then measuring the levels of LacZ in cells transfected with the two plasmids. Translation is presumed to start at the 11th AUG in the wild-type context but must start at the 12th AUG in the altered vector. Nonetheless, levels of LacZ were roughly the same in both cases (data not shown).

LacZ-CAT vectors. Unselected coexpression of the bacterial *lacZ* and *cat* genes was tested by expressing the upstream *lacZ* gene from the RSV LTR as a Gag-LacZ fusion protein and expressing the downstream *cat* gene in each of several ways: from a spliced subgenomic message (pLZSAC), from an internal RSV promoter (pLZUC), from an internal CMV promoter (pLZCC), or from EMCV IRES fragments (Fig. 1). Because our primary aim was to test the efficacy of the EMCV IRES strategy, and because the distance between the pyrimidine stretch in the IRES and the initiation codon has been reported to be critical (24), we constructed four different IRES-containing vectors (pLZIC1, pLZIC2, pLZIC3, and pLZΔIC). In the case of pLZIC1, the *cat* gene from pSV2CAT (15) was inserted at the new *Hin*dIII site of the EMCV IRES fragment (see above). In this vector, a 29-bp sequence of the CAT 5′ nontranslated region separates the CAT initiation codon from the 3′ end of the EMCV IRES. To assess the effect of this separation on CAT expression, we also converted the CAT initiation codon to an *Nco*I site and joined it to the *Nco*I site of either the altered EMCV IRES (pLZIC2) or the wild-type EMCV IRES (pLZIC3). Finally, plasmid pLZΔIC was constructed as a control for internal initiation, with the *cat* gene placed at the upstream *Hin*dIII site of the EMCV fragment. The EMCV fragment was only 232 bp long in this case and did not include the entire IRES sequence.

QT6 cells were transfected with each plasmid to confirm the integrity of the two genes. Cell extracts were prepared 40 h after transfection, and CAT activity relative to LacZ activity was measured (Table 1). All of the vectors expressed LacZ, and all except pLZΔIC expressed CAT, but the ratios of CAT to LacZ varied among them. The relative levels of CAT expression were roughly twofold higher when the *cat* gene was placed directly at the 3′ end of the EMCV IRES (pLZIC2 and -3) than when it was separated by a 29-bp untranslated fragment (pLZIC1) or expressed from the RSV promoter (pLZUC). The highest levels of CAT expression were obtained from the CMV promoter, which was previously shown to be more powerful than the RSV promoter in mammalian expression vectors (10). CAT expression from a

TABLE 1. Gene expression LacZ-CAT vectors

Vector	Relative CAT activity (%)[a]	
	In transfected cells	In infected cells
pLZΔIC	0.6	0.75
pLZIC1	49, 48	46, 62
pLZIC2	89, 84	75, 101
pLZUC	26, 58	68
pLZCC	274, 468	56, 74
pLZSAC	5, 6	4, 4

[a] For each vector, the amount of CAT activity was determined per unit of LacZ activity in the same extract. These values were then expressed as percentages of the values from pLZIC3-transfected or -infected cells. Numbers from two independent experiments are represented. CAT activities in pLZΔIC-infected or -transfected cells were similar to those of mock-transfected or -infected cells.

spliced message was only 5 to 10% that of the IRES vectors. There was no CAT expression from the 232-bp EMCV fragment of pLZΔIC, which confirms the requirement of an IRES for internal initiation. Thus, the ratio of CAT to LacZ activities depended on the mechanism used to drive expression of CAT.

To test for virus particle production and for expression of the *lacZ* and *cat* genes from a proviral genome, virus particles were generated by cotransfection of QT6 cells with a helper plasmid and one of the retroviral plasmids. Media were collected 40 h after transfection and used to infect fresh QT6 cells. After another 48 h, the infected cells were assayed in three ways. First, we estimated the number of LacZ infectious units per microgram of transfected DNA (LacZ titers) by staining with X-Gal and counting the resulting blue cells. LacZ titers were consistently higher from the IRES vectors than from the other proviruses. For example, in five experiments in which pLZIC1 and the U3 internal promoter vector were directly compared, the titer of the double-promoter vector was 23% ± 4% (mean ± standard error of the mean) that of the IRES vector. Thus, even if LacZ titers underestimate true titers of the double-promoter vector by twofold (see below), the IRES vectors still produce higher levels of viral particles. Similarly, in a single experiment, the LacZ titers of pLZSAC and pLZCC were found to be 35 and 40% that of pLZIC1, respectively. These results are consistent with the possibilities that negative interactions between promoters partially inhibit the production of genomic RNA from the LTR (8, 9) and that expression of spliced subgenomic mRNA is at the expense of full-length genomic RNA (see below).

Second, we estimated the amount of CAT expression relative to LacZ expression by preparing cell extracts from the infected cells and assaying them for the two enzymes. As for the transfectants, more CAT than LacZ was expressed from vLZIC2 and vLZIC3 (v denotes the proviral state) than from vLXIC1, vLZUC, or vLZSAC (Table 1). Surprisingly, the ratio of CAT to LacZ was severalfold lower for vLZCC than for pLZCC. This difference may reflect an effect of the integrative state of the provirus, although this was not observed in mammalian cells (10). For the other six vectors, however, the relative expression of the two genes was not greatly affected by the state of integration.

The biochemical assays of infected cells showed that the levels of CAT relative to LacZ from the double-promoter and IRES vectors varied less than twofold. They did not, however, address the issue of whether both genes are reliably coexpressed within individual cells. To test for coexpression, we performed a third set of assays in which we doubly stained infected cultures with antibodies to LacZ and CAT. Cells from infected cultures were seeded on coverslips 40 h after infection and grown for another 24 h. The cells were then fixed and stained with rhodamine-tagged anti-CAT and fluorescein-tagged anti-LacZ. As shown in Fig. 2 and Table 2, the extent of coexpression varied greatly among vectors. More than 90% of the vLZIC1-, -2-, and -3-infected cells that were LacZ positive were also CAT positive. In contrast, in vLZUC-infected cultures, only 6% of infected cells were stained for both CAT and LacZ; two-thirds of the infected cells made detectable levels of only the upstream *lacZ* gene, and one-quarter expressed only the downstream *cat* gene. This result suggests that the two promoters in the vector interfered with each other. To determine whether the interfering effect of the two promoters occurred only when the two promoters in the vector were identical, we substituted the internal U3 fragment with a fragment harboring an immediate-early CMV promoter. In this case, expression of LacZ and CAT was also noncoordinate in most (60%) of the infected cells; the other 40% stained positively for both LacZ and CAT. There was little CAT staining detectable above background in the vLZSAC-infected cells and none in the vLZΔIC-infected cells. The low levels of CAT activity measured by enzymatic assays in cells infected with vLZSAC presumably limited detection by immunostaining. Thus, IRES vectors direct more reliable coexpression of LacZ and CAT than do double-promoter or regulated-splicing vectors.

LacZ–v-Src vectors. To test the general utility of the IRES approach for expressing two genes from a proviral genome, we constructed a second set of vectors in which the viral oncogene v-*src* was placed downstream of the *lacZ* gene (Fig. 1). The downstream v-*src* gene was expressed from a spliced subgenomic mRNA (pLZSAS), from an internal RSV promoter (pLZUS), or from the EMCV IRES fragment (pLZIS). In all vectors, LacZ was expressed as a Gag-LacZ fusion protein from the 5′ LTR, and gene expression was estimated by visual inspection of immunohistochemically stained infected cells.

Virus particles were generated by cotransfection of QT6 cells with a helper plasmid and a LacZ–v-Src vector. Media were harvested 40 h after transfection and used to infect fresh QT6 cells. The vector pLZ20, which expresses LacZ alone, was used as a control. The infected cells were stained with X-Gal 48 h after infection to estimate LacZ titers. LacZ titers from four independent experiments indicated that the LacZ titers of pLZUS and pLZSAS were 8% ± 3% and 16% ± 10% that of pLZIS. This order is the same as that observed with the LacZ-CAT vectors and supports the idea that production of subgenomic RNA occurs at the expense of genomic RNA. The titers of vLZIS were 78, 125, and 169% those of vLZ20 in three experiments, indicating that insertion of the IRES does not interfere with virus production.

Coexpression in cells infected by LacZ–v-Src vectors was assessed by immunofluorescence using fluorescein-tagged anti-LacZ and Texas red-tagged anti-v-Src. Background staining by anti-v-Src was relatively high in all cultures, possibly because of cross-reaction of the antibody with endogenous c-Src. Nonetheless, specific staining attributable to v-Src was detectable, and it was evident that proviruses expressing v-Src from the EMCV IRES showed higher rates of coexpression than did the other two LacZ–v-Src proviruses (Fig. 3). Coexpression of LacZ and v-Src was

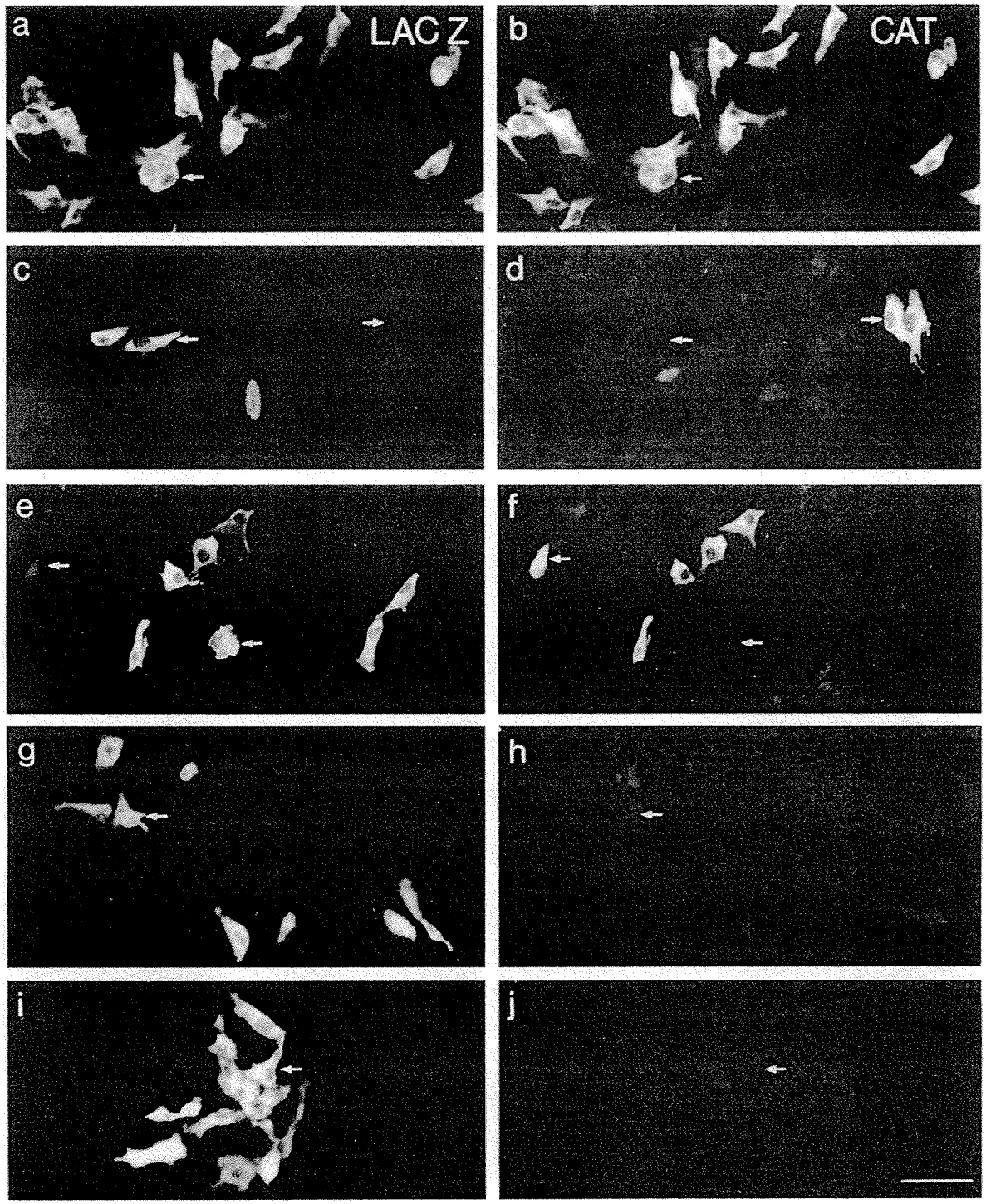


FIG. 2. Coexpression of LacZ and CAT from proviral genomes. Cultures of QT6 cells that had been infected with LacZ-CAT vectors were doubly stained with antibodies to LacZ and to CAT as described in Materials and Methods. Fields were then photographed with fluorescein optics to show LacZ (a, c, e, g, and i) and with rhodamine optics to show CAT (b, d, f, h, and j). Vectors were the IRES vector, pLZIC1 (a and b); the internal U3 promoter vector, pLZUC (c and d); the internal CMV promoter, pLZCC (e and f); the regulated splicing vector, pLZSAC (g and h); and the truncated IRES vector, pLZΔIC (i and j). LacZ and CAT are reliably coexpressed only in pLZIC1-infected cells. Arrows mark corresponding points on the two micrographs of each pair. In all cultures, <5% of the cells were antibody stained; unstained cells are only barely visible in these micrographs. The bar represents 50 μm.

TABLE 2. Coexpression of LacZ and CAT in QT6 cells infected with LacZ-CAT vectors[a]

Vector	% of infected cells expressing:		
	CAT only	LacZ only	LacZ and CAT
pLZΔIC	0	100	0
pLZIC1	3	0	97
pLZIC2	7	3	90
pLZIC3	2	2	96
pLZUC	25	69	6
pLZCC	34	26	40
pLZSAC	3	90	7

[a] Cells were doubly stained for CAT and LacZ as described in Materials and Methods. In each case, 200 stained cells were counted from randomly chosen fields.

detected in more than 90% of the vLZIS-infected cells but in less than 10% of the vLZUS- and vLZSAS-infected cells. Most of the vLZUS-infected cells that expressed LacZ stained weakly if at all above the background level with the v-Src antibody, and most of the v-Src-positive cells were negative for LacZ expression. In the case of vLZSAS-infected cells, there were roughly threefold more LacZ-positive than v-Src-positive cells, although again weak expression of v-Src might have been obscured by the endogenous immunoreactive material. Thus, LacZ and v-Src were more reliably coexpressed from the IRES vector than from the double-promoter or regulated-splicing vectors.

In examining the immunostained cultures, a second indication of v-Src expression became apparent: most v-Src-positive cells were rounded and refractile, as has been noted previously with fibroblasts infected or transfected with v-Src expression vectors (30). In vLZIS-infected cultures, most of the LacZ-positive cells detected immunohistochemically dis-

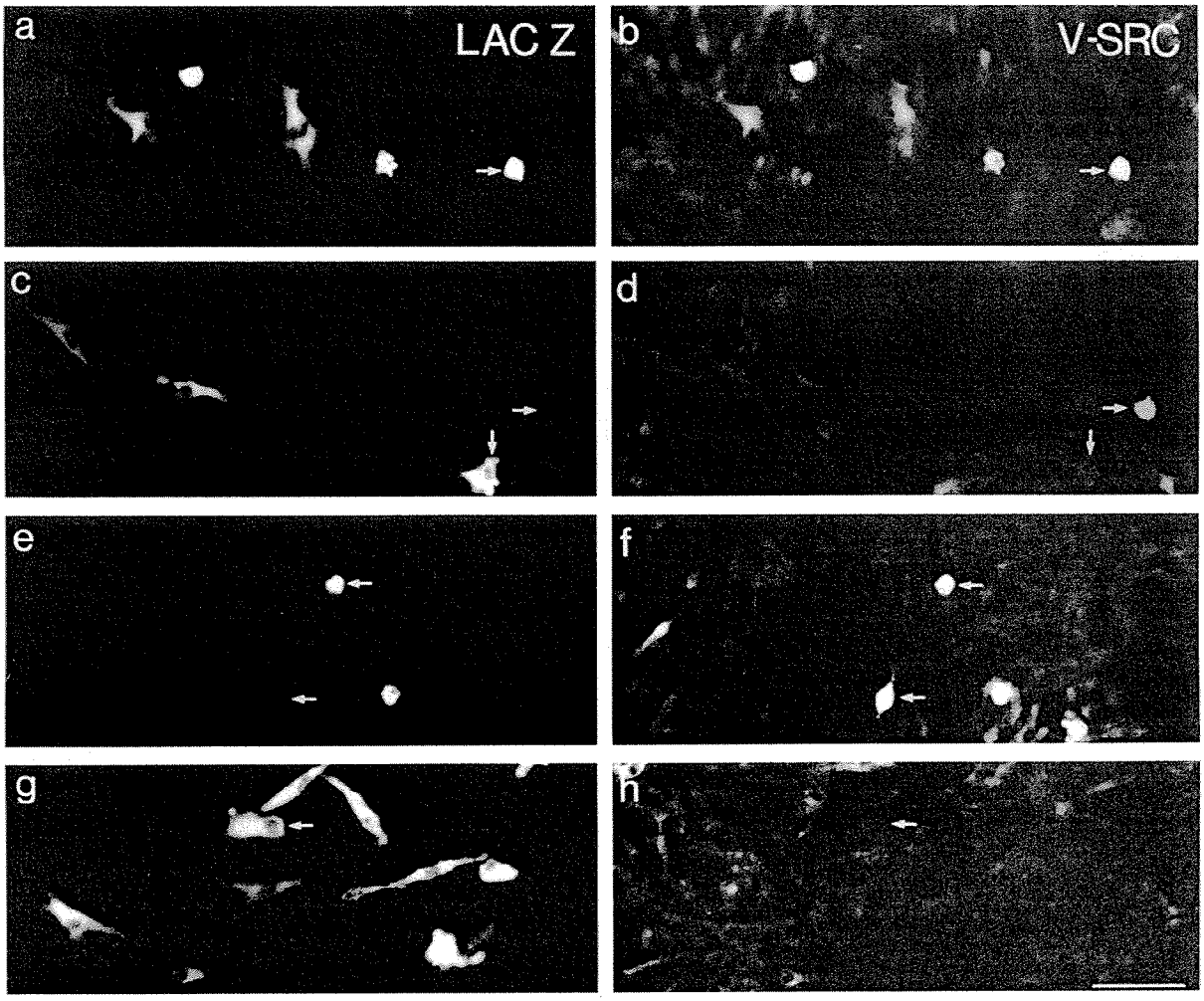


FIG. 3. Coexpression of LacZ and v-Src from proviral genomes. Cultures of QT6 cells that had been infected with LacZ–v-Src vectors were doubly stained with antibodies to LacZ and to v-Src as described in Materials and Methods. Fields were then photographed with fluorescein optics to show LacZ (a, c, e, and g) and with rhodamine to show v-Src (b, d, f, and h). Vectors were the IRES vector, pLZIS (a and b); the internal U3 promoter vector, pLZUS (c and d); the regulated-splicing vector, pLZSAS (e and f); and the LacZ-only vector, pLZ20 (g and h). LacZ and v-Src are reliably coexpressed only from pLZIS-infected cells. Arrows mark corresponding points on the two micrographs of each pair. The bar represents 50 μm.

played this altered phenotype, whereas most LacZ-positive cells were morphologically normal in vLZUS- and vLZSAS-infected cultures. This result supports the conclusion that vLZIS is the best of the three vectors tested at coexpressing LacZ and v-Src.

LacZ-Neo^r vector. As a third test of the IRES strategy, we constructed two vectors in which the *lacZ* and *neo* genes were separated by the EMCV IRES. In the first, the *lacZ* gene was expressed from the 5′ LTR as a Gag-LacZ fusion protein and the *neo* gene was placed downstream of the IRES sequence at the *Nco*I site (Fig. 1). In the second, the *neo* gene was expressed from the LTR as a Gag-Neo^r fusion protein and the *lacZ* gene was placed downstream of the IRES at the *Nco*I site. Virus was produced from both vectors and used to infect QT6 cells. Neo^r cells were selected by growth in G418 and stained for LacZ. More than 90% of the Neo^r cells were LacZ positive after infection with either vector. Thus, the IRES can be used to express LacZ as a downstream gene and to reliably coexpress a reporter gene and a selectable marker.

LacZ–v-Src expression in ovo. Because LacZ and v-Src were efficiently coexpressed from pLZIS in vitro, this vector was selected for tests in ovo. Virus concentrate, derived either from pLZIS or from the LacZ control vector, pLZ10, was injected into embryos at 55 to 70 h of incubation (Hamburger and Hamilton stages 13 to 17). The embryos were recovered, fixed, and stained for LacZ 3 to 6 days after injection (stages 30 to 34 are equivalent to embryonic days 6 to 8 [E6 to E8]).

In initial experiments, we analyzed LacZ-positive cells in the skin because it is simple, thin, and easily examined in whole mount. Clusters of closely spaced LacZ-positive cells were assumed to be clones on the basis of statistical and double-labeled studies that have been detailed elsewhere (13, 45). In vLZ10 (LacZ only)-injected embryos, typical clones were composed of flat, polygonal cells with large central nuclei and relatively scant cytoplasm (Fig. 4a and b); the average number of cells per clone was about 40 at stages 30 to 34. Plastic sections through such clones revealed that the marked cells were well integrated into the epidermis and identified them as typical cells of the basal and peridermal layers (Fig. 4c and d; see Fig. 7 in reference 45 for similar studies of mouse skin). In contrast, most cells in vLZIS (LacZ–v-Src)-marked clones were markedly abnormal: they were large and irregularly shaped and contained multiple LacZ-negative inclusions that might have been vacuoles or supernumerary nuclei generated by disordered cytokenesis (Fig. 4e and f). Sections through these clones showed that the LacZ-positive cells were poorly integrated into the epidermis (Fig. 4g) and frequently seemed to be delaminating from its apical surface (Fig. 4h). In addition the number of cells per clone was greatly reduced when v-Src was present; the average number of cells per vLZIS-marked clone, counting each potential syncytial aggregate as a single cell, was around five at the stages examined. Finally, it is important to note that the incidence of such abnormalities was high: in one set of vLZIS-injected embryos analyzed at stage 30, 14 of 15 clones of skin cells were markedly aberrant. In embryos analyzed at stage 34, about 25% of the clones contained groups of normal-appearing cells, either in isolation or intermingled with abnormal cells, suggesting that expression of v-Src may be reduced or rendered less toxic at later developmental stages (data not shown). Nonetheless, more than 90% of the clones analyzed at this stage contained one or more clearly aberrant cells. Thus, at least some progeny of most LacZ-positive vLZIS-marked epidermal cells infected in ovo expressed sufficient levels of v-Src to perturb normal developmental processes.

In addition to skin, we dissected and examined the central nervous system of vLZIS-injected embryos. Only a few clones of LacZ-positive cells were found within the brain or spinal cord, and these will be described elsewhere. In roughly 50% of the embryos, however, we found aggregates of LacZ-positive cells in association with the meninges that ensheath the brain and spinal cord and with the loose mesenchyme that overlies the meninges. Examples from the lumbar segments of the spinal cord and from the optic tectum of the mesencephalon are shown in Fig. 5. In both areas, the LacZ-positive cells were heterogeneous in size and shape but were not grossly abnormal in morphology. Most striking, however, was the large size of the clones: each cluster contained hundreds to thousands of LacZ-positive cells. No comparably large clones of meningeal or mesenchymal cells were seen in vLZ10-injected embryos (data not shown), nor have we seen such clones in previous studies of cell lineage in optic tectum (13, 16, 18) or spinal cord (29). From these results, we conclude that vLZIS promotes coexpression of LacZ and v-Src in mesenchymal tissue as well as in skin.

DISCUSSION

The aim of this work was to construct a recombinant retroviral vector which would allow efficient coexpression of two genes in single infected cells. We have shown that use of the EMCV IRES to produce a bicistronic message provides a more efficient means of expressing two genes within individual cells than does use of two promoters or a regulated-splicing mechanism.

Retroviral vectors containing internal promoters are often compromised by the problem of promoter interference, which is dependent on the strength of the promoter and the context in which it is placed. Cullen et al. (4) showed that expression of the preproinsulin gene from the 3′ LTR of the avian leukosis virus occurred only when the upstream LTR was inactivated. This phenomenon was called transcriptional interference and is reminescent of a more natural case observed in avian leukosis virus-induced bursal tumors (38, 41). Such tumors have been associated with insertion of the virus upstream of c-*myc*. In most of these tumors, the provirus was found to bear mutations near its 5′ end that were correlated with expression of c-*myc* from the 3′ LTR (14). Independently, Emmerman and Temin (8, 9) showed that in a double-promoter spleen necrosis virus vector, one gene is usually suppressed when there is selection for the other, regardless of the position of the genes relative to each other. This suppression was found to be epigenetic, reversible, and *cis* acting. It was attributed to a change in chromatin conformation near the active promoter, which in turn inhibited transcription from nearby promoters. More recently, Hoeben et al. (20) have suggested that suppression is associated with methylation and is dependent on the site of integration of the virus. This, however, was not observed by Emmerman and Temin (8).

Different problems occur when splice acceptor sites are used to construct double-expression vectors. If the aim were to produce maximal levels of both gene products, optimal expression would be reached when half of the RNA is spliced. Thus, each gene would be expressed at only 50% of the level expected if no splicing occurred. In addition, the amount of genomic RNA available for packaging, and potentially the viral titer, would be reduced by half. Finally,

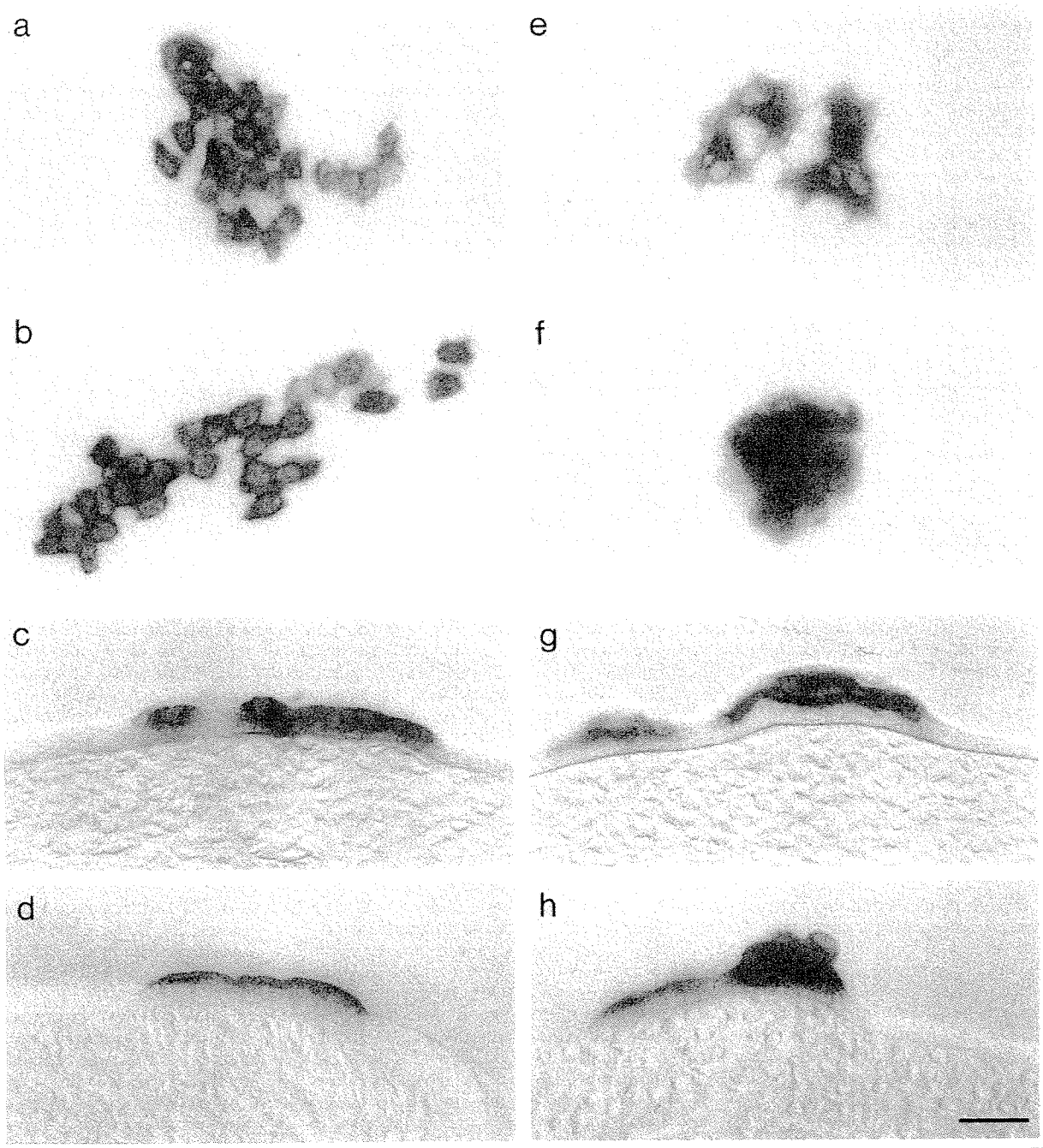


FIG. 4. Clones of LacZ-positive cells in the skin of chicken embryos that had been injected with vLZ10 (LacZ only; a to d) or vLZIS (LacZ–IRES–v-Src; e to h). Embryos were injected on E3 and examined on E6 to E8. (a, b, e, and f) Whole mounts, photographed with bright field optics; (c, d, g, and h) plastic sections photographed with Nomarski optics. Aberrant morphology of LacZ-positive, vLZIS-marked cells demonstrates that these cells express v-Src. The bar represents 50 μm.

regulation of alternative splicing is not well understood in any system, and it is already clear that splicing efficiency in RSV is context dependent and that discrete intronic sequences are responsible for the regulation of splicing (1, 2, 25, 48, 49). Thus, it is not currently possible to engineer vectors that produce spliced and unspliced RNAs in predictable ratios. In addition, for RSV the translational start site for *gag* is retained in the spliced message and could interfere with translation from the AUG of the downstream gene.

To obviate the problems encountered with the double-expression vectors that use internal promoters or regulated splicing, we tested the EMCV IRES. In the vectors that we

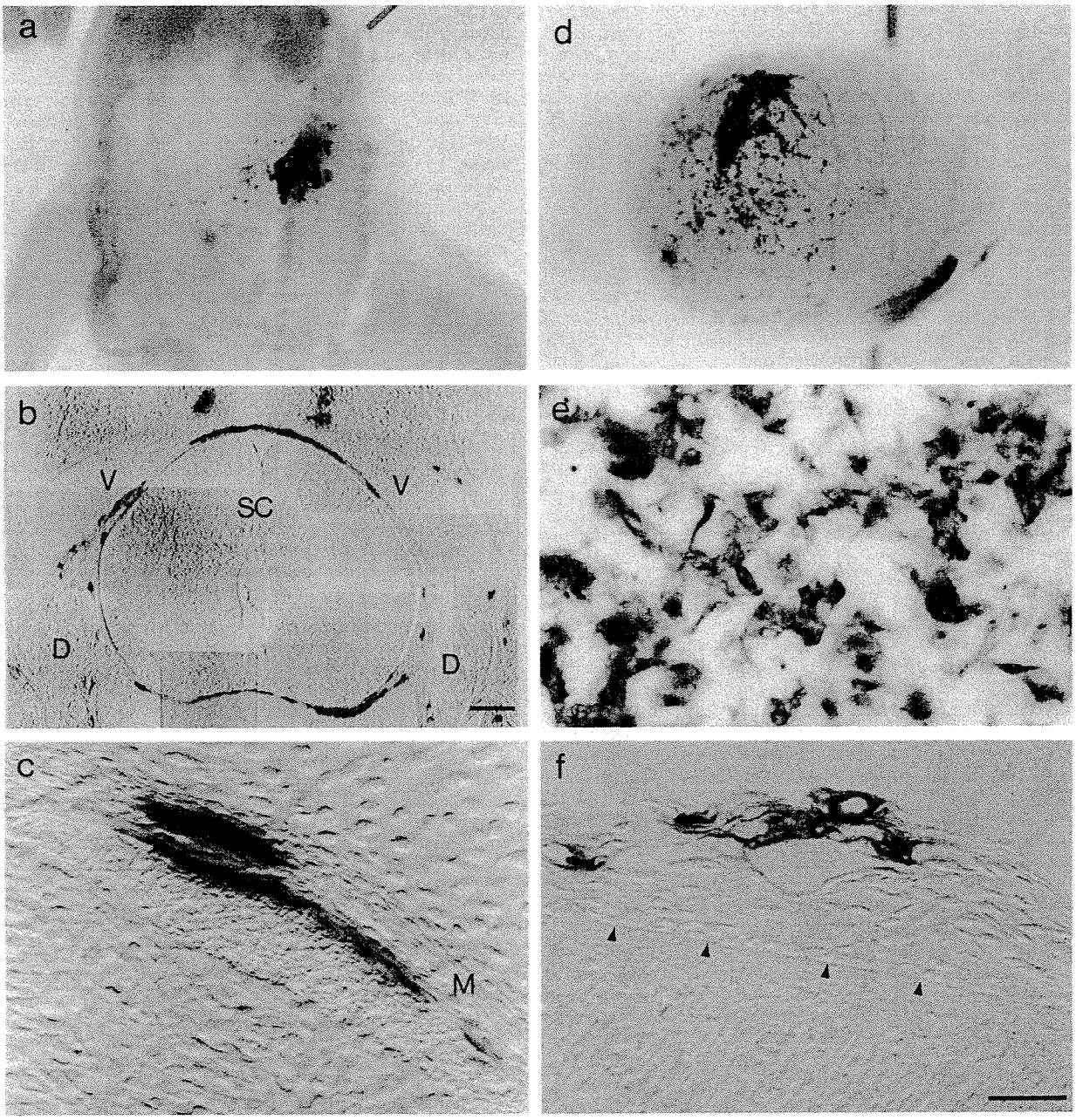


FIG. 5. Clones of LacZ-positive cells in the spinal (a to c) and tectal (d to f) meninges of chicken embryos that were injected with vLZIS on E3 and examined on E8. (a) Large clone of cells surrounding and extending laterally from the spinal cord. (b) Photomontage of a section through vertebral column shows that LacZ-positive cells ensheath but do not penetrate the spinal cord (SC) and extend into mesenchyme surrounding vertebral bodies (V) and dorsal root ganglia (D). Groups of LacZ-positive mesenchymal cells more distant from the cord were also present in this clone but are not visible in this field. (c) At higher power and with Nomarski optics, some of the perispinal cells are seen to be associated with the meninges (M) external to a fiber tract in the spinal cord. (d) Clone of LacZ-positive cells on the surface of the optic tectum. (e) At higher power in whole mount, the cells seem to be disordered, heterogeneous in shape, and numerous. (f) Section photographed with Nomarski optics showing LacZ-positive meningeal cells external to the apical surface (arrowheads) of the tectum. The bar represents 1 mm in panel a, 125 μm in panel b, and 50 μm in panel c, e, and f, and 0.5 mm in panel d.

constructed, both genes were expressed from the 5′ LTR of RSV and translation of the downstream gene in the dicistronic message occurred by ribosomes binding internally to the IRES. The efficiency of translation of the second cistron in such dicistronic mRNAs has been shown to be similar to that of the corresponding gene in a monocistronic message, indicating that interference does not occur at the translational level (23). Several lines of evidence presented here

indicate that the IRES provides the best available method for coexpressing two genes from a retroviral vector. (i) In a population of infected cells grown under no selection, more that 90% of cells infected with an IRES vector coexpressed LacZ and CAT, whereas most cells infected with an internal promoter vector expressed either one gene or the other. (ii) Similarly, coexpression of LacZ and v-Src in infected QT6 cells was also substantially more efficient from the IRES vector than from either an internal promoter or a regulated splicing vector. (iii) The IRES vector permits reliable coexpression of Neo^r and LacZ (>90%) in either order (Neo^r-IRES-LacZ or LacZ-IRES-Neo^r). (iv) The levels of expression of LacZ, CAT, and v-Src determined either by enzymatic assays or by visual inspection of immunostained cells were high. (v) For both LacZ-CAT and LacZ–v-Src vectors, LacZ titers were consistently higher with the IRES vectors. Together, these results provide strong evidence that the EMCV IRES is more efficient than either of the previously used approaches at coexpressing two genes from a recombinant provirus.

We chose the LacZ–IRES–v-Src vector (pLZIS) for injection into embryos because previous studies in vivo and in vitro (26, 30, 42, 46, 47), as well as our observations on cultured cells (Fig. 3), suggested that expression of v-Src would be likely to affect cellular morphology. Indeed, in both skin and meninges, vLZIS-marked cells were consistently different from vLZ10-marked cells in number or size, indicating that most LacZ-positive cells also expressed v-Src. In addition, although our aim was not to study the biology of v-Src, these experiments provide new data on the effects of v-Src on cells in complex tissue. First, expression of v-Src appears to block the normal proliferation of epidermal cells. This inhibition might result either from a general toxic effect of the oncogene or from a more specific interference with endogenous mechanisms of growth control; we cannot distinguish between these alternatives. Our results are somewhat different from those of Stoker et al. (46), who saw disturbances of growth control in only one of several clones of epidermal cells marked by a LacZ–v-Src vector. One possible explanation for the difference is that Stoker et al. used a regulated splicing vector in which many LacZ-positive cells (about 40%) did not express detectable levels of v-Src, and other cells may have expressed lower levels than occur in vLZIS-infected cells. In addition, whereas we analyzed clones at E6 to E8, Stoker et al. did not examine tissue until E12 to E14, by which time v-*src*-poisoned cells might have died and only unaffected cells survived. In contrast, expression of v-Src appears to cause loss of proliferative control in the mesenchyme and meninges. This result is consistent with the known ability of v-*src* to transform mesenchymal cells (e.g., fibroblasts) in vitro and to induce fibrosarcomas in vivo (42). The large clones of mesenchymal and meningeal cells that we observed might have eventually formed tumors, as has been observed previously following infection with replication-defective v-*src*-containing virus (47); however, we confined our observations to short times (3 to 6 days) after infection.

IRES-based retroviral vectors of the sort described in this report may be useful for a variety of purposes. (i) Vectors encoding the LacZ marker and a second bioactive gene can be used to generate small clones of transgenic cells in a wild-type environment. As shown here for vLZIS, the transgenic cells can be identified histochemically, and the effects of the second gene can be assessed by analyzing the phenotype of the labeled cells. (ii) The LacZ-Neo^r vector can be used in transcription activation studies: mutations are introduced in a promoter, Neo^r clones are isolated, and the effect of the mutation on promoter activity is measured by LacZ assays (13a). (iii) Similarly, if vectors for gene therapy were constructed by linking *neo* and the gene of interest by an IRES, selection for G418 resistance should ensure expression of the second gene. Finally, it is important to note that use of the IRES strategy described here is not limited to retroviral vectors. For example, transgenic mice expressing a marker gene (e.g., *lacZ*) and a bioactive gene from a single promoter would allow direct correlation of patterns of transgenic expression with patterns of transgenic effect.

ACKNOWLEDGMENTS

We thank Anne Palmenberg for pE5LVP0, Henry Huang for the CAT fragment with the altered initiation codon, Jeff Milbrandt for the CMV promoter fragment, Robin Morris-Valero and Jeanette Cunningham for technical assistance, Deni Galileo for advice on immunohistochemistry, and Grace Gray for helpful comments.

This work was supported by grants from the McKnight Foundation and NIH to J.R.S. and from NIH to J.E.M.

REFERENCES

1. **Arrigo, S., and K. Beemon.** 1988. Regulation of Rous sarcoma virus RNA splicing and stability. Mol. Cell. Biol. **8**:4858–4867.
2. **Berberich, S. L., and M. Stoltzfus.** 1991. Mutations in the region of the Rous sarcoma virus 3′ splice sites: implications for regulation of alternative splicing. J. Virol. **65**:2640–2646.
3. **Boshart, M., F. Weber, G. Jahn, K. Dorsch-Hasler, B. Fleckenstein, and W. Schaffner.** 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell **41**:521–530.
4. **Cullen, B. R., P. T. Lomedico, and G. Ju.** 1984. Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis. Nature (London) **307**:241–245.
5. **Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. Goodman.** 1983. Correction to the nucleotide sequence of the src gene of Rous sarcoma virus. Nature (London) **301**:736–739.
6. **DeLorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop.** 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**:50–61.
7. **Elroy-Stein, O., T. R. Fuerst, and B. Moss.** 1989. Cap-independent translation of mRNA conferred by encephalomyocarditis virus 5′ sequence improves the performance of the vaccinia virus/bacteriophage T7 hybrid expression system. Proc. Natl. Acad. Sci. USA **86**:6126–6130.
8. **Emmerman, M., and H. M. Temin.** 1984. Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. Cell **39**:459–467.
9. **Emmerman, M., and H. M. Temin.** 1986. Quantitative analysis of gene suppression in retrovirus vectors. Mol. Cell. Biol. **6**:792–800.
10. **Foecking, M. K., and H. Hofstetter.** 1986. Powerful and versatile enhancer-promoter unit for mammalian expression vectors. Gene **45**:101–105.
11. **Frank, E., and J. R. Sanes.** 1991. Lineage of neurons and glia in chick dorsal root ganglia: analysis *in vivo* with a recombinant retrovirus. Development **111**:895–908.
12. **Friedmann, T.** 1989. Progress toward human gene therapy. Science **244**:1275–1281.
13. **Galileo, D. S., G. E. Gray, G. C. Owens, J. Majors, and J. R. Sanes.** 1990. Neurons and glia arise from a common progenitor in chick optic tectum: demonstration with two retroviruses and cell type-specific antibodies. Proc. Natl. Acad. Sci. USA **87**:458–462.

13a. **Ghattas, I. R., and J. E. Majors.** Unpublished data.

14. **Goodenow, M. M., and W. S. Hayward.** 1987. 5′ long terminal repeats of *myc*-associated proviruses appear structurally intact

but are functionally impaired in tumors induced by avian leukosis virus. J. Virol. 61:2489–2498.
15. Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2:1044–1051.
16. Gray, G. E., J. C. Glover, J. Majors, and J. Sanes. 1988. Radial arrangement of clonally related cells in the chicken optic tectum: lineage analysis with a recombinant retrovirus. Proc. Natl. Acad. Sci. USA 85:7356–7360.
17. Gray, G. E., S. M. Leber, and J. R. Sanes. 1990. Migratory pattern of clonally related cells in the developing central nervous system of the chick embryo. Experientia 46:929–940.
18. Gray, G. E., and J. R. Sanes. 1991. Migratory paths and phenotypic choices of clonally related cells in the avian optic tectum. Neuron 6:211–225.
19. Hamburger, V., and H. L. Hamilton. 1951. A series of normal stages in the development of the chick embryo. J. Morphol. 88:49–92.
20. Hoeben, R. C., A. A. J. Migchielsen, R. C. M. van der Jagt, H. van Ormondt, and A. J. van der Eb. 1991. Inactivation of the Moloney murine leukemia virus long terminal repeat in murine fibroblast cell lines is associated with methylation and dependent on its chromosomal position. J. Virol. 65:904–912.
21. Jakobovits, E. B., J. E. Majors, and H. E. Varmus. 1984. Hormonal regulation of the Rous sarcoma virus *src* gene via a heterologous promoter defines a threshold dose for cellular transformation. Cell 38:757–765.
22. Jang, S. K., M. V. Davies, R. J. Kaufman, and E. Wimmer. 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5′ nontranslated region of encephalomyocarditis virus RNA in vivo. J. Virol. 63:1651–1660.
23. Jang, S. K., H.-G. Krausslich, M. J. H. Nicklin, G. M. Duke, A. C. Palmenberg, and E. Wimmer. 1988. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J. Virol. 62:2636–2643.
24. Jang, S. K., and E. Wimmer. 1990. Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosome entry site and involvement of a cellular 57-kd RNA binding protein. Genes Dev. 4:1560–1572.
25. Katz, R. A., M. Kotler, and A. M. Skalka. 1988. *cis*-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implication for mechanisms of control. J. Virol. 62: 2686–2695.
26. Keller, G., and E. F. Wagner. 1989. Expression of v-src induces a myeloproliferative disease in bone marrow-reconstituted mice. Genes Dev. 3:827–837.
27. Kozak, M. 1989. The scanning model for translation: an update. J. Cell Biol. 108:229–241.
28. Kriegler, M. 1990. Gene transfer and expression: a laboratory manual. Stockton Press, New York.
29. Leber, S. M., S. M. Breedlove, and J. R. Sanes. 1990. Lineage, arrangement, and death of clonally related motoneurons in spinal cord. J. Neurosci. 10:2451–2462.
30. Levy, J. B., and J. S. Brugge. 1989. Biological and biochemical properties of the c-*src*$^+$ gene product overexpressed in chicken embryo fibroblasts. Mol. Cell. Biol. 9:3332–3341.
31. Lopata, M. A., D. W. Cleveland, and B. Sollner-Webb. 1984. High level transient expression of a chloramphenicol acetyl transferase gene by DEAE-dextran mediated DNA transfection coupled with a dimethyl sulfoxide or glycerol shock treatment. Nucleic Acids Res. 12:5707–5717.
32. Majors, J., and H. E. Varmus. 1983. A small region of the mouse mammary tumor virus long terminal repeat confers glucocorticoid hormone regulation on a linked heterologous gene. Proc. Natl. Acad. Sci. USA 80:5866–5870.
33. Maness, P. 1986. pp60$^{c\text{-}src}$ encoded by the proto-oncogene c-*src* is a product of sensory neurons. J. Neurosci. Res. 16:127–139.
34. Miller, A. D., and G. J. Rosman. 1989. Improved retroviral vectors for gene transfer and expression. BioTechniques 9:980–990.
35. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. Morinaga, Y., T. Franceschini, S. Inouye, and M. Inouye. 1984. Improvement of oligonucleotide-directed site-specific mutagenesis using double-stranded plasmid DNA. Bio/Technology 2:636–639.
37. Moscovici, C., G. Moscovici, H. Jimenez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt. 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell 11:95–103.
38. Neel, B. G., W. S. Hayward, H. L. Robinson, J. Fang, and S. M. Astrin. 1981. Avian leukosis virus-mediated tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell 23:323–334.
39. Overell, R. W., K. E. Weisser, and D. Cosman. 1988. Stably transmitted triple-promoter retroviral vectors and their use in transformation of primary mammalian cells. Mol. Cell. Biol. 8:1803–1808.
40. Parks, G. D., G. M. Duke, and A. C. Palmenberg. 1986. Encephalomyocarditis virus 3C protease: efficient cell-free expression from clones which link viral 5′ noncoding sequences to the P3 region. J. Virol. 60:376–384.
41. Payne, G. S., S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, and H. E. Varmus. 1981. Analysis of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for maintenance of the tumor state. Cell 23:311–322.
42. Pimental, E. 1989. Oncogenes, 2nd ed., vol. 2, p. 115–158. CRC Press, Boca Raton, Fla.
43. Price, J., D. Turner, and C. Cepko. 1987. Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. Proc. Natl. Acad. Sci. USA 84:156–160.
44. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. Sanes, J. R., J. L. R. Rubenstein, and J.-F. Nicolas. 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. EMBO J. 5:3133–3142.
46. Stoker, A. W., C. Hatier, and M. Bissell. 1990. The embryonic environment strongly attenuated v-src oncogenesis in mesenchymal and epithelial tissues, but not in endothelia. J. Cell Biol. 111:217–228.
47. Stoker, A. W., and M. H. Sieweke. 1989. v-src induces clonal sarcomas and rapid metastasis following transduction with a replication-defective retrovirus. Proc. Natl. Acad. Sci. USA 86:10123–10127.
48. Stoltzfus, C. M., and S. J. Fogarty. 1989. Multiple regions in the Rous sarcoma virus *src* gene intron act in *cis* to affect the accumulation of unspliced RNA. J. Virol. 63:1669–1676.
49. Stoltzfus, C. M., S. K. Lorenzen, and S. L. Berberich. 1987. Noncoding region between the *env* and *src* genes of Rous sarcoma virus influences splicing efficiency at the *src* gene 3′ splice site. J. Virol. 61:177–184.
50. Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus. 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. USA 78:124–128.
51. Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop. 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. USA 80:2519–2523.
52. Temin, H. M. 1986. Retrovirus vectors for gene transfer: efficient integration into and expression of exogenous DNA in vertebrate cell genomes, p. 149–187. *In* R. Kucherlapati (ed.), Gene transfer. Plenum Press, New York.
53. van der Eb, A. J., and F. L. Graham. 1980. Assay of transforming activity of tumor virus DNA. Methods Enzymol. 65:826–839.
54. Weiss, R., N. Teich, H. E. Varmus, and J. Coffin (ed.). 1982. RNA tumor viruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Nucleic Acids Research, Vol. 20, No. 6 1293–1299

Retroviral vectors containing putative internal ribosome entry sites: development of a polycistronic gene transfer system and applications to human gene therapy

Richard A.Morgan, Larry Couture, Orna Elroy-Stein[1], Jack Ragheb, Bernard Moss[1] and W.French Anderson

Molecular Hematology Branch, National Heart, Lung, and Blood Institute and [1]Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

Received December 13, 1991; Revised and Accepted February 10, 1992

ABSTRACT

Recombinant retroviral vectors producing multicistronic mRNAs were constructed. Picornavirus putative internal ribosome entry sites (IRES) were used to confer cap-independent translation of an internal cistron. Internal cistrons were engineered by ligation of various lengths of the IRES of encephalomyocarditis (EMC) virus or polio virus to the E. coli chloramphenicol acetyltransferase (CAT) gene. The IRES/CAT fusions were introduced into retroviral vectors 3′ to the translation stop codon of the neomycin phosphotransferase (NEO) gene, and the molecular constructs transfected into retroviral vector packaging lines. Retroviral vector producer cells efficiently express the internal CAT gene product only when the full length IRES is used. Both the EMC/CAT and polio/CAT retroviral vectors produced high titer vector supernatant capable of productive transduction of target cells. To test the generality of this gene transfer system, a retroviral vector containing an IRES fusion to the human adenosine deaminase (ADA) gene was constructed. Producer cell supernatant was used to transduce NIH/3T3 cells, and transduced cells were shown to express NEO, and ADA. Novel three-gene-containing retroviral vectors were constructed by introducing the EMC/ADA fusion into either an existing internal-promoter-containing vector, or a polio/CAT bicistronic vector. Producer cell clones of the three-gene vectors synthesize all three gene products, were of high titer, and could productively transduce NIH/3T3 cells. By utilizing cap-independent translation units, IRES vectors can produce polycistronic mRNAs which enhance the ability of retroviral-mediated gene transfer to engineer cells to produce multiple foreign proteins.

INTRODUCTION

Retroviral-mediated gene transfer is a highly efficient method of gene transfer that has recently seen its first clinical applications (1). Several types of retroviral vectors have been constructed that use different mechanisms for achieving gene expression. The most common vector designs use the long terminal repeat (LTR) of the retrovirus backbone and, an internal promoter sequence, to promote gene expression (2). Still other manipulations have been used to produce complex splicing vectors which use splice donor/acceptor sequences to generate multiple mRNAs (3,4). One potential problem with retroviral vectors containing multiple transcription units is that if selection is applied for one gene, expression of the other gene can be reduced or lost completely, (this is termed promoter suppression, see references 5,6). To try to avoid promoter suppression and to construct vectors that readily express multiple genes, it may be necessary to exploit alternate expression systems. Gene expression in the picornaviridae family of viruses is unusual in that their 5′ mRNA terminus is pUp... and they possess long untranslated leader sequences (7,8). Analysis of picornavirus gene expression has produced a consistent body of work that suggests that picornaviruses are able to bypass the standard ribosome scanning model of translation and begin translation at internal sites (9–12). Putative internal ribosome entry sites (IRES) have been identified in the long 5′ untranslated regions of picornaviruses (these sequences have also been termed ribosome landing pads, RLP). These IRES elements can be removed from their viral setting, and linked to unrelated genes to produce polycistronic mRNAs (9,11,13). Initial reports describing the application of these elements in retroviral-mediated gene transfer have recently appeared (14,15).

Herein we describe that picornavirus IRES elements can be linked to various genes and that the fusions, when inserted into retroviral vectors, are translated to yield functional gene products. In addition, we extend previous reports (14,15) by producing several three-gene-containing retroviral vectors. These IRES vectors permit multiple proteins to be produced from a single vector without alternative splicing or multiple transcriptions units and avoid the potential of promoter suppression. Furthermore, the coupling of translation of two (or more) different proteins may have significant applications in human gene therapy where the expression, in a given cell, of multiple heterologous proteins or distinct subunits of a multimeric protein is necessary.

MATERIAL AND METHODS

Molecular Constructs

EMC/CAT vectors, were constructed from the T7 RNA polymerase expression plasmid pOS6 (also referred to as pT7EMCAT, 16−18). Cla I plus Bsp MII were used to excise the T7-EMC/CAT expression cassette. The resulting fragment was made blunt ended by Klenow fill-in, and ligated into the Hind III cut/Klenow fill-in site of pG1N to produce pG1NECt and pG1NECt-R. EMC deletions mutants were constructed by digesting pOS6 with Bsp MII plus Apa I, Kpn I, or Nco I, followed by cloning into pG1N (as above) to yield pG1NECt-Δ200, pG1ENCt-Δ525, and pG1NCt, respectively. To facilitate further manipulations, the EMC IRES was isolated from pOS6 using polymerase chain reaction (PCR) amplification/restriction enzyme site addition (oligonucleotide primers, 5'-AACGGTTT-CCCTCGAGCGGGATCA-3' plus 5'-TTTGTTAGCAGCCGG-ATCGT-3') yielding a fragment with Xho I ends which was cloned into the Xho I site of Bluescript II KS+(Stratagene, La Jolla CA) to produce pEMC-F. PCR was similarly used to produce a fragments containing the ADA gene using the SAX retroviral vector (19) as a template (oligonucleotide primers 5'-TGCGAGACCATGGGACAGACGCCC-3' plus 5'-CGG-AAGTGTGATCACCTAGGCGAC-3'). The ADA fragment was digested with Nco I, cloned into the Nco I plus Sma I sites in pEMC-F to produce pEMCADA. The EMC/ADA fragment was excised by Sst I digestion/T4 DNA polymerase fill-in plus Xho I digestion and ligated to Apa I cut/T4 DNA polymerase fill-in plus Xho I cut retroviral vector pG1N, to produce pG1NEA. The starting vectors for the triple gene constructs, LSCSN and LNSvCt were produced by inserting the soluble CD4 gene and CAT gene into the Eco RI plus Xho I (for CD4) or Hind III (for CAT) sites of LXSN and LNSX respectively (20). The EMC-ADA fragment was excised from pEMCADA by Xba I digestion/Klenow fill-in plus Xho I digestion and ligated to Bam HI cut/Klenow fill-in plus Xho I LSCSN to produce LSCEASN. To produce LNEASCt, pEMCADA digested with Xho I plus Sst I, filled in with Klenow and ligated to Bam HI cut/Klenow fill-in LNSvCt. The polio IRES element was isolated by PCR amplification/restriction site addition (oligonucleotide primers 5'-CCCAGATCTCCACGTGGCGGC-3' plus 5'-ACCGGAA-GGCCTATCCAATTC-3') using pPV16 as a template (21). PCR generated a fragment with Bgl II and Stu I ends which was ligated into Bam HI plus Stu I cut LNSvCt to yield LNPCt. LNEAPCt and LNPCtEA were produced by inserting the EMC/ADA fragment from pEMCADA (Sst I plus Xho I with T4 DNA polymerase fill-in) into the Nru I site or Cla I/Klenow fill-in site of LNPCt respectively. LCtSN was constructed by ligating a Hind HIII cut/Klenow fill-in CAT fragment into the Hpa I site of LXSN (the LXSN vector used in this report has had the normal Moloney U3 promoter region removed and substituted with the U3 region from Harvey murine sarcoma virus) The vectors pG1, pG1N2, and pG1N, are similar to the LN vector (20), but with additional cloning sites (G1 vectors kindly provided by Dr. Paul Tolstoshev, Genetic Therapy Inc. Gaithersburg, MD).

Cell Culture and Vector production

Retroviral vector producer cell lines were generated by the micro-ping-pong procedure (22,23). In brief, 50μg of DNA was used to transfect (via calcium phosphate coprecipitation) a mixture of the ecotropic packaging cell line GP+E-86 (24), and the amphotropic packaging cell line PA317 (25). The packaging cell line mixtures are maintained in culture for at least one week to permit vector amplification. Selection for vector integration is obtained by growth in the presence of the neomycin analog G418 (400 μg/ml active concentration). Transductions of mouse NIH/3T3 cells and Mink lung fibroblasts (ATCC CCL 64) were conducted by incubation of cells with recombinant viral supernatant (MOI = 1)containing 8μg/ml polybrene at 37° for 2hr, followed by removal of virus-containing medium and replacement with fresh culture medium. Transduced cell populations were selected by growth in G418 (400μg/ml) for 10−14 days. Cell clones were obtained using cloning rings following limiting dilution.

Gene Expression Assays

CAT enzyme assays were performed by first lysing cells (at 4°C) in 0.25M Tris-HCl(pH 7.5)/0.1% NP-40, followed by freezing on dry ice, thawing at 37°C (5 min), heating to 60°C (15 min) and removal of cellular debris by centrifugation (top speed, eppendorf microcentrifuge, 4°C, 5 min). After normalization for equal amounts of protein, cell extracts were mixed with acetyl-CoEnzyme A and ^{14}C-chloramphenicol and incubated at 37°C for 1−4 hr. as necessary to stay within the linear range of CAT activity. Chloramphenicol and acetylated products were extracted with ethyl acetate and applied to thin layer chromatography plates. Chromatographs were run in 95% $CHCl_3$, 5% methanol. Imaging was obtained by autoradiography and quantitation by direct beta particle counting of the TLC plates on a Betascope 603 instrument. Southern blot analysis was performed by subjecting restriction enzyme digested DNA samples to agarose gel electrophoresis, transfer to nylon membranes with UV cross-linking, and hybridization with a radiolabeled probe. Northern blot analysis was performed on formaldehyde agarose gels using RNA extracted with RNazol (CINNA Biotecx, Friendswood TX) and selected for poly A containing sequences by oligo-dT linked magnetic beads (Dynal Co. Great Neck, NY). ADA assays were performed on starch gels as described (26) and relative ADA actively was determined by scanning the resultant photographs on a laser densitometer and calculating the ratio of the areas of human ADA to mouse ADA enzymes. Soluble CD4 levels were measured using a CD4/gp120 capture ELISA (American Biotechnologies, Cambridge MA). NEO gene activity was measured using a NPT II ELISA (5 Prime, 3 Prime Inc. West Chester, PA).

RESULTS

Construction of CAT IRES vectors

To determine if the picornavirus IRES elements could function in a retroviral vector we constructed a series of CAT reporter gene vectors using the IRES elements from both the EMC and polio viruses. As starting material we used the EMC/CAT gene fusion from the plasmid pOS6 (16). The full length EMC/CAT fusion was excised and transferred in both orientations into the retroviral vector G1N to generating G1NECt (sense orientation) and G1NECt-R (reverse orientation). The distance between the NEO stop codon and the CAT start codon in the G1NECt construct is approximately 700 bp, and contains 9 intervening start codons plus 19 interrupting stop codons in all three reading frames. Next we made three deletions of the EMC IRES sequence and transferred these into the same retroviral backbone. In G1NECt-Δ200, G1NECt-Δ525, and G1NCt we deleted respectively; 200 bp, 525 bp, or all of the sequences between

the 5′ end of the EMC IRES leader sequence and the start of the CAT gene. These constructs along with a control CAT vector were then introduced into retroviral vector packaging cell lines and the coculture expanded for two weeks to allow vector spread. Southern blot analysis indicated that equivalent amounts of each vector were present in these lines (data not shown). Cell lysates were prepared and equal amounts of protein used to assay for CAT activities as described (see methods).

Figure 1A, shows significant CAT activity from the G1NECt IRES vector in comparison to the activity driven by the strong chimeric LTR present in LCtSN. Quantitation of CAT activity, determined in the linear range of the assay, indicated that G1NECt containing cells produce 72% of the LCtSN activity. To rule out the possibility that the EMC IRES was serving as a promoter element in the context of a retroviral vector, the construct G1NECt-R, with the EMC/CAT fusion in the reverse orientation, was produced and tested. No CAT activity was observed from the reverse orientation EMC/CAT vector. Analysis of the EMC deletion mutants indicate that efficient expression of the internal CAT gene is dependent on the presence of a full length IRES element. Deletion of 200 bp from the EMC IRES decreases activity of G1NECt-Δ200 to 4% of the full length EMC construct. The amount of CAT activity gradually increases as the EMC CAT start codon is brought closer to the NEO stop codon, with CAT activity increasing to 6% of control for G1NECt-Δ525, and 18% of control for G1NCt.

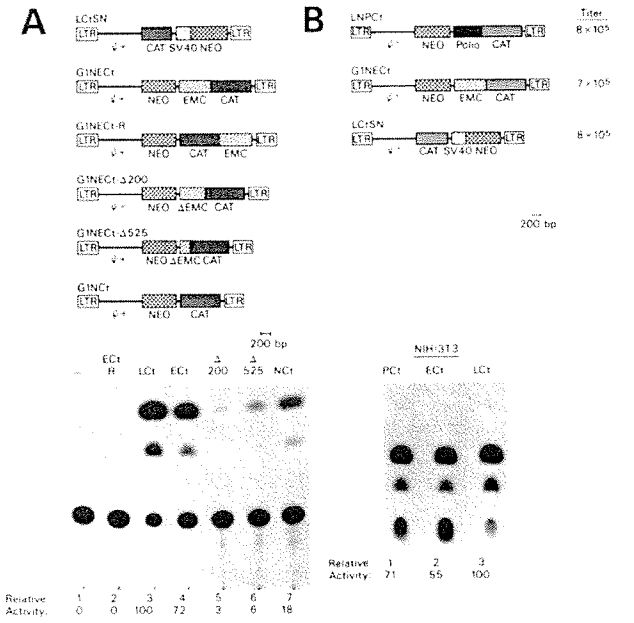


Figure 1. IRES CAT Vectors. Shown on the top of the figure are diagrams of the indicated IRES CAT vectors. Below is shown the autoradiograms from CAT enzyme analysis (1hr incubation). Panel A, lane 1, producer cells transfected with pOS6; lane 2, G1NECt-R; lane 3, LCtSN; lane 4, G1NECt; lane 5, G1NECt-Δ200; lane 6, G1NECt-Δ525; lane 7, G1NCt. In panel B, LNPCt, G1NECt, and LCtSN retroviral vector-containing supernatant (titer of producer cells populations are indicated to the right of each vector diagram) was used to transduce NIH/3T3 cells. After selection in G418 containing medium cell lysates were prepared for CAT assays; lane 1, LNPCt transduced NIH/3T3 cells; lane 2, G1NECt; lane 3, LCtSN. Relative activity was calculated from the mean of at least three independent assays (all samples within linear range) where the percent conversion of the LCtSN control was set to 100%.

In the next series of experiments, we isolated the IRES from poliovirus and used it to construct a retroviral vector. PCR was used to generate a fragment containing the IRES element from the 5′ untranslated region of poliovirus (Mahoney strain). The polio IRES was inserted 3′to the NEO stop codon and upstream of a CAT reporter gene to generate LNPCt. This vector along with the EMC IRES construct (G1NECt), and the positive control vector (LCtSN) were transfected into packaging cell cocultures. The cultures were grown for one week in standard medium and selected for vector containing cells by growth for two weeks in G418-containing culture medium. Completely selected cultures were used to harvest retroviral-vector-containing supernatant for NEO^R titer determination. The titer for all three vectors was similar when assayed on NIH/3T3 cells (approximately 7.5×10^5 $G418^R$ cfu/ml, fig 1B). Retroviral vector supernatant from the G1NECt LNPCt, and LCtSN producer cells was used to transduce NIH/3T3 cells. Following transduction, the cells were cultured for 10 days in the presence of G418. After selection, lysates were prepared for CAT assays. The representative CAT activity for each transduction is shown in fig 1B. The data indicate that both the G1NECt and LNPCt IRES vectors can produce functional retroviral vector particles that productively transfer and express an IRES/reporter gene in target cells (LNPCt 70%, and G1NECt 55% of LCtSN).

Construction of an EMC human ADA vector

To evaluate the use of IRES elements in the construction of retroviral vectors for potential human gene therapy applications, a fusion between the EMC IRES and the human adenosine deaminase (ADA) gene was assembled and introduced into a

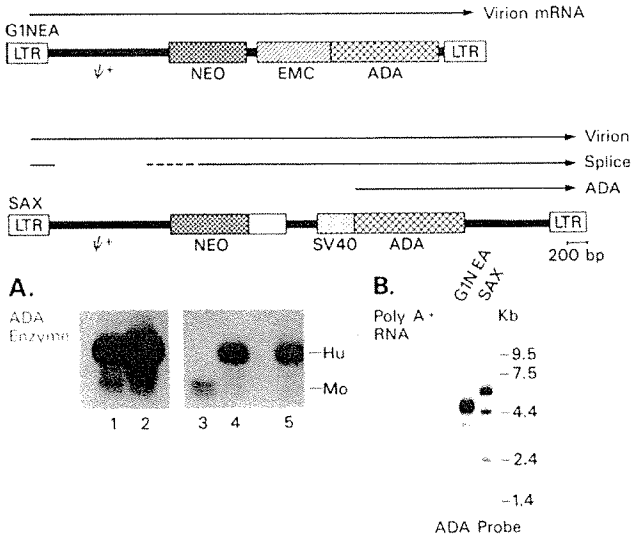


Figure 2. ADA IRES Vector. Shown on the top of the figure is a diagram of the EMC/ADA vector G1NEA, and the control ADA vector, SAX. Arrows indicate location of mRNA species, the exact location of the SAX splice acceptor site has not been mapped (dashed line). Panel A, starch gel analysis for ADA enzyme activity, equal amounts of total cell lysates were used for each sample, the location of the human (Hu) and mouse (Mo) ADA enzymes are indicated. Lane 1, SAX producer cells; lane 2, G1NEA producer cells; lane 3, NIH/3T3 cells; lane 4, SAX transduced 3T3 cells; and lane 5, G1NEA transduced 3T3 cells. Panel B, Northern blot analysis using 5μg of poly A+ mRNA hybridized with a human ADA probe. The origin of each sample is indicated on the top of the lanes.

retroviral vector. To do this, a DNA fragment containing the human ADA gene was synthesized, using PCR, and cloned into the pEMC-F plasmid to generate pEMCADA. The EMC/ADA fusion was excised from pEMCADA and inserted into the retroviral vector G1N yielding G1NEA (Fig. 2, top).

DNA for the G1NEA vector and the control ADA vector SAX were transfected into packaging cell line cocultures. The cocultures were grown for 1 week in standard culture medium, then selected for stable vector integration by culture for 2 weeks in the presence of the neomycin analog G418. The G418 selected producer cell populations were used to generate vector containing supernatant for titer determinations, and were subjected to gene expression analysis.

Figure 2, panel A shows the results of ADA starch gel analysis on the G1NEA producer cells (lane 2) and SAX control producer cells (lane 1). Both producer cell populations made large amounts of human ADA. Retroviral-vector-containing supernatant from the producer cell populations were then used to transduce NIH/3T3 cells and to determine the vector titer on 3T3 cells. Both producer cell populations yielded high titer vector supernatants with SAX being 1.9×10^6 G418^{R}cfu/ml and G1NEA being 1.2×10^6 G418^{R}cfu/ml. The G418^R 3T3 cells were next assayed for ADA activity. ADA starch gel analysis demonstrated functional transfer of the human ADA gene into the 3T3 cells by the G1NEA IRES vector (Fig. 2, panel A, lane 5). Northern blot analysis (Fig. 2, panel B), was used to visualize the RNA transcripts from the two vectors in transduced 3T3 cells. For SAX, as previously reported (19), a full length LTR transcript as well as the internal SV40 transcript and a spliced subgenomic transcript were detected with the ADA probe. In the case of G1NEA, only one full length transcript is identified by Northern blot analysis with the ADA probe (a very small amount of what could be a spliced transcript may also be seen in RNA from the G1NEA cells).

Construction of triple gene vectors

To test the versatility of IRES elements in the construction of complex retroviral vectors, we inserted the EMC/ADA fusion gene into two different double-gene vectors to generate three-gene-containing vectors. The first recipient vector, LSCSN, uses the LTR to promote the expression of the anti-HIV agent soluble CD4 (sCD4) and an internal SV40 early region promoter to drive the NEO selectable marker gene (27). The EMC/ADA fragment was introduced after the sCD4 translation stop codon and 5′ to the start of the SV40 promoter to generate LSCEASN (Fig. 3). LSCEASN DNA was transfected into a packaging cell line coculture that was grown for one week before being passaged, at limiting dilution, into G418-containing medium. Twelve G418^R producer cell clones were isolated and expanded for analysis. All 12 G418^R producer cell clones synthesize both the human ADA enzyme and the sCD4 protein (fig. 3).

The second two-gene retroviral vector used as recipient for the EMC/ADA fragment was LNSCt, a vector that uses the LTR to drive NEO expression and has an internal SV40 promoter directing CAT expression. EMC/ADA was inserted 3′ to the NEO gene stop codon and upstream of the SV40 promoter to generate LNEASCt (Fig. 4). LNEASCt DNA was transfected into packaging cells, cultured for one week, and G418^R producer cell clones were isolated by limiting dilution. Twelve producer cell clones were expanded and used to isolate vector containing supernatant to determine G418^R titer, and analyzed for both CAT and ADA gene expression. Figure 4 shows that all twelve producer cell clones had both CAT (panel A) and human ADA (panel B) enzyme activity. The titer from the twelve clones showed a typical distribution, ranging from 4×10^4 G418^{R}cfu/ml for clone 10 to 4×10^6 G418^{R}cfu/ml for clone 4.

Retroviral vector-containing supernatant from LNEASCt producer cell clone 12 was used to transduce NIH/3T3 cells. Six G418^R 3T3 cells clones were isolated by limiting dilution, and expanded for analysis. To verify the integrity of the integrated provirus, genomic DNA was isolated from each clone digested with restriction enzyme Sst I, and the digested DNA subjected to Southern blot analysis using a NEO probe (Fig 5, Southern). All six clones produced the predicted 5593 bp band with no apparent rearrangements or deletions (note, the images seen below the proviral band were the result of contamination of the intensifying screen with previous CAT assay samples). Next, RNA was isolated and subjected to formaldehyde gel

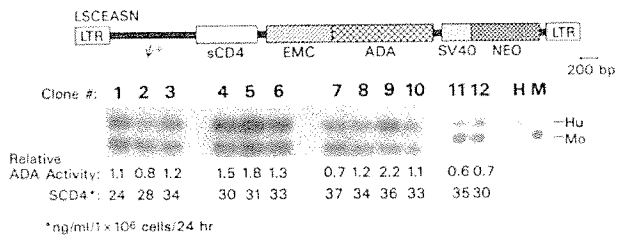


Figure 3. sCD4-ADA-NEO Triple-Gene Vector. Shown on the top of the figure is a diagram of the LSCEASN (sCD4, ADA, and NEO) triple-gene vector. Below is shown the results of ADA starch gel analysis and sCD4 ELISA from 12 G418^R producer cell clones (numbers 1–12, H=human control, M=mouse control). Relative ADA activity is determined as the ratio of the intensity of the human to mouse ADA enzyme bands while the amounts of sCD4 produced in the culture medium was determined using standards supplied by the manufacturer.

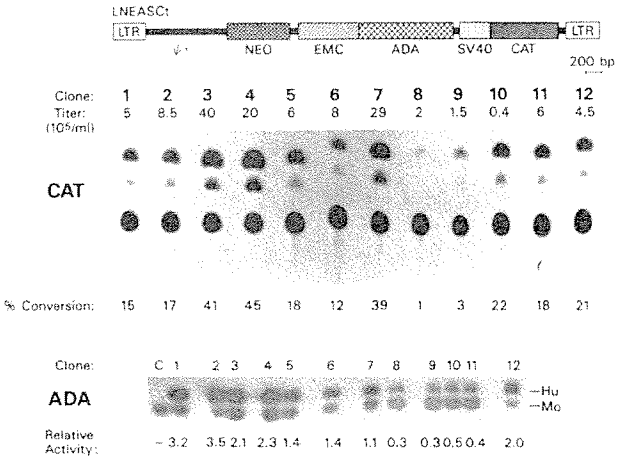


Figure 4. NEO-ADA-CAT Triple-Gene Vector. Shown on the top of the figure is a diagram of the LNEASCt (NEO, ADA, and CAT) triple-gene vector. Shown first below, autoradiogram of resultant CAT activity (1hr incubation, % conversion as indicated) from 12 G418^R producer cell clones (numbers 1–12). The titer, measured on NIH/3T3 cells, of G418^{R}cfu/ml is indicated below the producer cell clone number. Lower panel, ADA starch gel analysis from the 12 producer cell clones (numbers 1–12), C = NIH/3T3 cells, human (Hu) and mouse Mo) ADA bands are indicated. Relative ADA activity is determined as the ratio of the intensity of the human to mouse ADA enzyme bands.

electrophoresis/Northern blot analysis with a ADA probe (fig 5, Northern). All six clones produced a single RNA transcript of the proper length using an ADA probe for detection. Enzyme assays were then performed to measure CAT and ADA with NEO being measured by an NPT II ELISA. ADA and CAT activity was documented in 6 of 6 3T3 clones and further, NPT II protein was also detected in all clones (Fig. 5, ADA, CAT, NPT II).

For the last series of vectors, we inserted the EMC/ADA gene fragment into the bicistronic LNPCt vector. The EMC/ADA fragment was inserted 5′ to the polio IRES to produce LNEAPCt or 3′ to the CAT gene to produce LNPCtEA (Fig. 6). Both DNA's were transfected into producer cell cocultures along with the control bicistronic vectors LNPCt and G1NEA. The cultures were grown for one week, then selected for vector-containing cells by growth in G418 containing medium for an additional 7 days. Retroviral-vector-containing supernatant from the G418^R producer cells was collected and cells harvested for DNA and RNA isolation. The supernatant was used for titer determination on NIH/3T3 cells and to transduce a mink lung cell line.

The stability of the LNEAPCt and LNPCtEA vectors was analyzed by subjecting producer cell DNA to digestion with the Sst I restriction enzyme (Sst I cuts once in each LTR), followed by Southern blot analysis (Fig. 6, Southern). Each vector produced a band of the expected size of 5860bp, with no sign of rearranged species. Northern blot analysis, additionally revealed no significant species of RNA's other than those predicted to originate from a single LTR promoted transcript (Fig. 6, Northern). Analysis of vector-containing-supernatant on NIH/3T3 cells yielded titers of 1×10^5 G418^R cfu/ml for LNPCtEA, 3×10^5 G418^R cfu/ml for LNEAPCt, 6×10^5 G418^R cfu/ml for LNPCt, and 8×10^5 G418^R cfu/ml for G1NEA. The same supernatants were used to transduce mink lung cells that were then selected for vector-containing cells by growth in G418 containing medium. Mink cells were then harvested for ADA and CAT enzyme analysis (Fig. 6, ADA and CAT). The results of these assays indicate that each of the tricistronic vectors was fully functional, they each produced all three gene products, ADA, CAT, and were G418^R (ie. NEO^R producing). For each enzyme tested, the bicistronic vector yielded more activity than

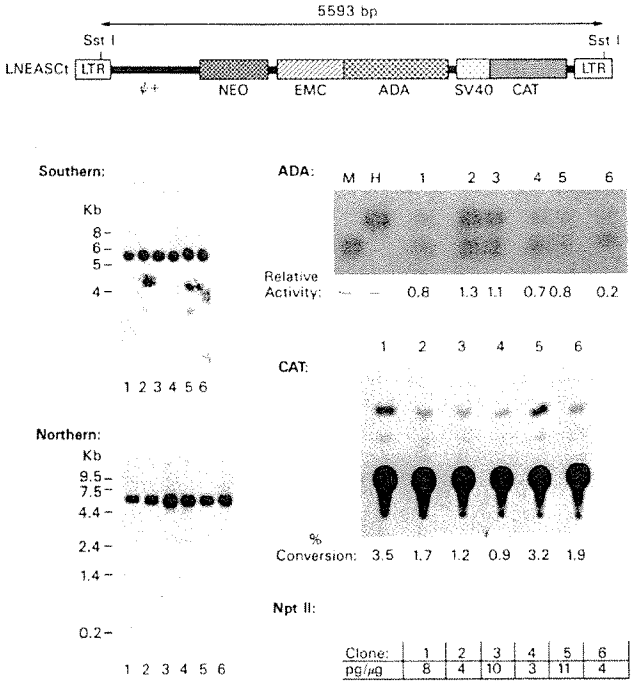


Figure 5. Analysis Triple-Gene Transduced Cells. Shown on the top of the figure is a diagram of the LNEASCt (NEO, ADA, and CAT) triple-gene vector. Southern, 20μg of genomic DNA from 6 independent NIH/3T3 cell clones was digested with Sst I and subjected to Southern blot analysis using a human ADA probe (note, the signals below the proviral band were the result of contamination of the intensifying screen with ^{14}C from a previous CAT assay). Northern, 20μg of total cell RNA from the 6 3T3 clones was subjected to Northern blot analysis using a human ADA probe. ADA, starch gel analysis of 6 3T3 cell clones, relative ADA activity is determined as the ratio of the intensity of the human to mouse ADA enzyme bands. CAT, CAT enzyme analysis (4hr incubation, % conversion as indicated) from 6 3T3 cell clones. Npt II, data from NPT II ELISA using cell lysates from 6 3T3 cell clones (expressed as pg NPT II/μg total protein). The numbering of each lane corresponds to the particular NIH/3T3 cell clone used.

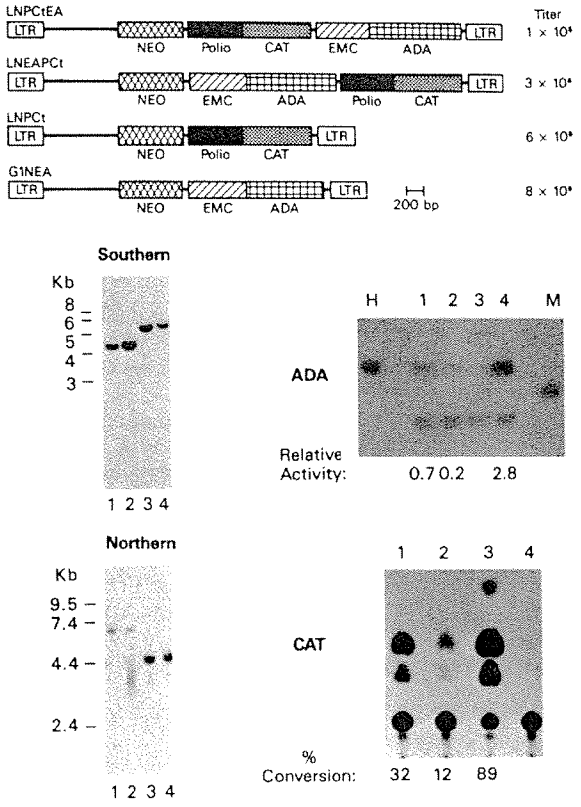


Figure 6. Analysis of Tricistronic Vectors. Shown on the top of the figure are diagrams of each of the individual vectors used along with the resultant titer of the producer cells, G418^R cfu/ml measured on NIH/3T3 cells. Southern, 20μg of genomic DNA from producer cells LNPCt (lane 1), G1NEA, (lane 2), LNEAPCt (lane 3), and LNPCtEA (lane 4) was digested with Sst I and subject to Southern blot analysis using an NEO probe for detection. Northern, 20μg of total RNA from producer cells LNEAPCt (lane 1), LNPCtEA (lane 2), LNPCt (lane 3), and G1NEA (lane 4) was subjected to Northern blot analysis using a human ADA probe for transcript visualization. ADA, starch gel analysis of mink lung cells transduced with the following vectors; LNEAPCt (lane 1), LNPCtEA (lane 2), LNPCt (lane 3), and G1NEA (lane 4). Lane H is a human ADA control, lane M is a mouse cell control (the mink ADA enzyme migrates below the mouse control). Relative ADA activity is determined as the ratio of the intensity of the human to mink ADA enzyme bands. CAT, CAT enzyme assays (% conversion as indicated) from mink lung cells transduced with the following vectors; LNEAPCt (lane 1), LNPCtEA (lane 2), LNPCt (lane 3), and G1NEA (lane 4).

the corresponding tricistronic vector, with LNEAPCt yielding both greater CAT and ADA activity than LNPCtEA (see discussion).

DISCUSSION

We describe retroviral vectors in which the expression of an internal protein coding sequence is facilitated by the use of picornavirus-derived putative internal ribosome entry sites (IRES vectors). As test constructs, we assembled retroviral vectors in which the polio virus or EMC virus IRES was linked to the prokaryotic reporter gene CAT (fig. 1). Both of these constructs yielded functional retroviral vectors that were able to transfer productively the reporter genes to NIH/3T3 cells. The relative amount of reporter gene enzyme activity was approximately 50% – 75% of that generated by direct transcription/translation using the strong promoter element in LCtSN (the retroviral LTR in LCtSN is derived from Harvey murine sarcoma virus and is 2 – 3×more active than the standard Moloney virus LTR used in the IRES vectors, unpublished observations).

To verify that the expression of the internal CAT gene was the result of putative cap-independent translation mediated by the IRES element, we constructed a series of control vectors (fig. 1). First, the opposite orientation vector, G1NECt-R, displays no CAT activity suggesting that EMC sequences are not functioning as general promoters. Second, deletion mutants indicate that removing 200bp from the 5′ end of the EMC IRES element results in near abolishment of CAT activity. Third, in the G1NCt vector, deletion of all of the IRES element, (leaving no interrupting start or stop codons), results in the CAT start codon being brought to within 20 bp of the NEO stop codon. This vector produces 18% of control CAT activity versus 72% for the complete EMC IRES. These data demonstrate that insertion of the full length IRES element greatly facilitates second gene translation. The mechanism for this is likely to be the same employed by picornaviruses, that is, internal ribosome binding and subsequent cap-independent translation.

The next construct produced was devised to mimic retroviral vector designs that are currently being used in a human gene therapy protocol to treat adenosine deaminase deficiency (28). Here, the EMC IRES was joined to the human ADA gene, and the fusion inserted into a retroviral vector down stream from the stop codon of the NEO gene (fig. 2). The G1NEA vector was compared to the well described SAX ADA vector and found to behave similarly to SAX with respect to titer (1.9×10^6 cfu/ml for SAX, 1.2×10^6 cfu/ml for G1NEA) and ADA expression (see fig. 2, panels A and B). The results of ADA starch gel analysis indicate that EMC mediated expression of the internal ADA gene is approximately equivalent to that generated in the SAX vector. Northern blot analysis failed to demonstrate significant amounts of subgenomic ADA transcripts suggesting that the IRES is facilitating internal translation initiation.

In the IRES vectors G1NECt, LNPCt, and G1NEA a putative bicistronic mRNA is produced by transcription initiating in the 5′ LTR promoter. The retroviral LTR is a strong promoter and has been shown to be more active than internal promoters in vitro (4,29) and simple LTR vectors have been successfully utilized in vivo (30 – 32). While retroviral splicing vectors can also achieve the translation of two gene products using the LTR promoter element, it has been difficult to predict the efficiency of splicing in these vectors and activation of cryptic splice donor sequences can result in deletion of vector sequences (4,33).

Many retroviral vectors have been constructed that express two independent genes using an internal promoter element to drive the expression of the second gene (2). IRES vectors may prove to be an improvement over standard two-promoter vectors. Because only one promoter is used in IRES vectors, these vectors should avoid the phenomenon of promoter suppression (5,6). Furthermore, reporter gene expression from an internal promoter can be unstable (in long term culture), with or without selection for the LTR driven gene (34). To date, the LTR promoted bicistronic IRES vectors have been stable in culture (eight week culture period, data not shown).

We extended the use of IRES elements in retroviral vector design to assemble three-gene retroviral vectors (figs. 3, 4, and 6). Triple-gene IRES vectors avoid the use of multiple promoter elements (again avoiding possible promoter suppression) and are compatible with the generation of high titer functional retroviral vectors. The construction of triple-promoter-containing retroviral vectors has been previously reported (35). In this report, significant differences in the levels of RNA transcripts were observed. The LTR, SV40, and HSV tk promoters used in this previous report, yielded relative RNA levels of 42:6:1 respectively. This observed disparity in RNA levels could be avoided in IRES vectors.

We assembled two types of three-gene-containing vectors by inserting the EMC/ADA fusion into either existing internal promoter vectors (Fig. 3 and 4), or into the bicistronic LNPCt vector (Fig. 6). All arrangements of genes and promoters in the various three-gene vectors were capable of generating high titer stable producer cells which can productively transduce target cells. Further analysis of the various three-gene-containing vectors will be required for quantitative comparisons to assess which vector arrangement is best (work in progress). In comparison of the reporter gene expression produced from the bicistronic versus tricistronic vectors, greater expression of both the CAT or ADA gene is observed using the bicistronic system. The nature of this observation (either differences in RNA stability, translatability or both) is not know at this time. The observed difference between the two tricistronic vectors (LNEAPCt yielding more CAT and ADA than LNPCtEA) may be related to the distance between the two IRES elements (approximately 1300 bp in LNEAPCt versus 800 bp in LNPCtEA).

In multiple IRES constructs, a single transcription event generates a polycistronic mRNA that can be translated to yield multiple proteins. The coupling of independent protein translations can have several advantages in human gene therapy situations where multiple protein subunits are needed (eg. immunoglobulins, or T-cell receptors) or in systems where two heterologous proteins may be more efficient at a given task (eg. by combining multiple anti-HIV proteins in the same vector). It may further be advantageous to have the production of a potentially physiologically dangerous protein be coupled to that of a conditional cell lethal protein (eg. tumor necrosis factor with herpes virus thymidine kinase). It is these types of applications where IRES vectors may find their greatest usefulness.

ACKNOWLEDGEMENTS

We would like to thank Craig Lassy, James Mc Ardle and Kim Hiatt (GTI) for invaluable technical assistance and David Dichek for critical reading of the manuscript.

REFERENCES

1. Rosenberg,S.A, Aebersold, P., Kasid, A., Morgan, R.A., Cornetta., Karson,E., Lotze, M.T., Yang, J.C.m Toplain, S., Moen, R., Culver, L., Blaese, R.M., and Anderson, W.F. (1990) N. Engl. J. Med. **323**, 570–578.
2. McLachlin,J.R., Cornetta,K., Eglitis,M.E., and Anderson, W.F. (1990) Prog. Nucleic Acid Res. Mol. Biol. **32**, 91–135.
3. Cepko, C.L., Roberts, B.E., and Mulligan, R.C. (1984). Cell **37**, 1053–1062.
4. Bowtell,D.D.L., Cory,S., Johnson,G.R., and Gonda,T.J. (1988). J. Virol. **62**, 2464–2473.
5. Emerman, M., and Temin, H.M. (1984). Cell **39**, 459–467.
6. Emerman, M., and Temin,H.M. (1986). Mol. Cell. Biol. **6**, 792–800.
7. Hewelett,M.J., Rose,J.K., and Baltimore,D. (1976). Proc. Natl. Acad. Sci. USA **73**, 327–330.
8. Nomoto,A., Lee,Y.F., and Wimmer,E. (1976). Proc. Natl. Acad. Sci. USA. **73**, 375–380.
9. Pelletier, J., and Sonenberg, N. (1988). Nature, **334**, 320–325.
10. Jang,S.K., Krausslich,H.G., Nicklin,M.J., Duke,G.M., Palmenberg,A.C., and Wimmer,E. (1988). J. Virol. **62**, 2636–2643.
11. Pelletier, J., and Sonenberg, N. (1989). J. Virol. **63**, 441–444.
12. Jang, S.K., and Wimmer, E. (1990). Genes and Development. **4**, 1560–1572.
13. Jang, S.K., Davies,M.V., Kaufman,R.J., and Wimmer,E. (1989). J. Virol. **63**, 1651–1660.
14. Adam,M.A., Ramesh,N., Miller,A.D., and Osborne, W.R.A. (1991) J. Virol. **65**, 4985–4990.
15. Ghattas,I.R., Sanes,J.R., and Majors,J.E. (1991) Mol. Cell. Biol. **11**, 5848–5859.
16. Elroy-Stein, O., Fuerst, T.R., and Moss, B. (1989). Proc. Natl. Acad. USA. **86**, 6126–6130.
17. Elroy-Stein, O., and Moss, B. (1990). Proc. Natl. Acad. Sci. USA. **87**, 6743–6747.
18. Moss,B., Elroy-Stein,O., Mizukami,T., Alexander,W.A., and Fuerst,T.R. (1990). Nature. **348** 91–92.
19. Kantoff,P.E., Kohn,D.B., Mitsuya,H., Egilitis,M.E., McLachlin,J.R., Gilboa, E., Blaese,R.M., and Anderson,W.F. (1986). Proc. Natl. Acad. Sci. USA. **83**, 6563–6567.
20. Miller, A.D., and Rosman, G.J. (1989). Biotechniques. **7**, 980–990.
21. Van-Der Werf,S., Bradley,B., Wimmer,E., Studier,F.W., and Dunn,J.J. (1986) Proc. Natl. Acad. Sci. USA. **83**, 2330–2334.
22. Bestwick,R.K., Kozak,S.L., and Kabat,D. (1988). Proc. Natl. Acad. Sci. USA. **85**, 5404–5408.
23. Muenchau,D.D., Freeman,S.M., Cornetts, K., Zwiebel,J.A., and Anderson, W.F. (1990) Virology. **176**, 262–265.
24. Markowitz,D., Goff,S., and Bank,A. (1988). J. Virol. **62**, 1120–1124.
25. Miller, A.D., and Buttimore, C. (1986). Mol. Cell. Biol. **6**, 2895–2902.
26. Lim,B., Williams,D.A., and Orkin,S.H. (1987). Mol. Cell. Biol. **7**, 3458–3465.
27. Morgan,R.A., Muenchau, D.D., Looney,D., Gallo,R.C., and Anderson,W.F. (1990). AIDS Res. and Human Retroviruses. **6**, 183–191.
28. Anderson,W.F. (1990). Human Gene Therapy. **1**, 331–362.
29. Hock,R.A., Miller,A.D., and Osborne,W.R.A. (1889). Blood. **74**, 876–881.
30. Dick,J.E.M., Magli,M.C., Huszar,D., Phillips,R.A., and Bernstein,A. (1985). Cell. **42**, 71–79.
31. Eglitis,M.E., Kantoff,P., Gilboa,E., and Anderson,W.F. (1985). Science. **230**, 1395–1398.
32. Belmont,J.W.J., MacGregor,G.R., Wager-Smith,K., Hawkins,D., Villalon,D., Chang,S.M.W., and Caskey,C.T. (1988). Mol. Cell. Biol. **8**, 5116–5125.
33. McIvor, R.S. (1990). Virology. **176**, 652–655.
34. Yee, J.-K. L.X., Wolff, J.A., and Friedmann, T. (1989). **171**, 331–341.
35. Overell,R.W., Weisser,K.E., and Cosman,D. (1988). Mol. Cell. Biol. **8**, 1803–1808.

JOURNAL OF VIROLOGY, Apr. 1993, p. 2142–2148
0022-538X/93/042142-07$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 67, No. 4

Characterization of a Bicistronic Retroviral Vector Composed of the Swine Vesicular Disease Virus Internal Ribosome Entry Site

BING-FANG CHEN,[1] LIH-HWA HWANG,[2]* AND DING-SHINN CHEN[2]

Graduate Institute of Microbiology, College of Medicine, National Taiwan University,[1] and Hepatitis Research Center, National Taiwan University Hospital,[2] Taipei, Taiwan, Republic of China

Received 12 October 1992/Accepted 11 January 1993

We cloned the 5′ nontranslated region (NTR) from the genome of swine vesicular disease virus (SVDV), a member of the family *Picornaviridae*, and used it to construct a bicistronic retroviral vector. The vector is characterized by coexpression of two genes from a single transcript. We found that inclusion of the 5′ NTR of SVDV did not negate the viral vector titer. Protein analysis indicated that the 5′ NTR could efficiently direct internal initiation, thus allowing the downstream gene to be translated. Translation of the internally initiated porcine growth hormone gene was about 30-fold less than that when the porcine growth hormone gene was at the upstream position in NIH 3T3 cells but was about equivalent to that in HeLa cells, implying that some cellular factors that stimulated internal initiation of the SVDV 5′ NTR are present in HeLa cells. However, in G418-selected clones, the Neo^r-encoding gene was expressed with equivalent efficiency either at a downstream position or at an upstream position in either NIH 3T3 or HeLa cells. Compared with the conventional double-gene vector or the U3-based vector, the bicistronic vector coexpressed two genes much more efficiently, owing to elimination of promoter interference. Furthermore, this type of vector infected and expressed the target genes efficiently in two primary cell lines, rat embryo and human skin fibroblast cells, which we tested. These experimental data suggest a better design for the retroviral vector and provide evidence that internal initiation of the SVDV 5′ NTR was stimulated cell specifically.

Retroviral vectors have been widely used in gene transfer, since they have broad host ranges (8, 32). In many situations it is essential to express two genes from a single proviral genome, one for the selectable marker to facilitate the isolation of a stable cell lineage and the other for the gene of interest. Conventional retroviral vectors include an exogenous promoter inserted at a site within the retroviral transcriptional unit. The upstream gene is expressed from the retroviral long terminal repeat (LTR), and the downstream gene is expressed from the internal promoter (9, 13). The problem with such a design is competitive interference between two promoters (9, 10, 24). As a result, one of these two genes may not be transcribed efficiently in individual isolates. Several approaches have been used to improve the vector design, such as the use of a self-inactivating vector (13) or a U3-based vector (4, 15). However, we have shown that the exogenous gene expressed from a double-gene vector, whichever is currently available, is never expressed as efficiently as from a single-gene vector (17).

Recently, the picornavirus 5′ nontranslated region (NTR) has been demonstrated to possess a specific secondary structure, thus allowing the ribosome to bind to it directly and initiate protein translation from a downstream AUG codon (see review in reference 19). This mechanism is different from the ribosomal scanning model that is considered valid for the translation of most eukaryotic cellular and viral mRNAs (20). By taking advantage of the property of the 5′ NTR of picornavirus, it is possible to construct a single-transcript type of retroviral vector which can simultaneously express two exogenous genes. In this way, only one transcript is expressed from the LTR promoter; thus, transcriptional interference of two active promoters can be avoided. Indeed, this has been documented in two reports demonstrating that insertion of the internal ribosome entry site (IRES) of encephalomyocarditis virus (EMCV) in a retroviral vector allows two cistrons to be expressed efficiently from a single transcript (1, 12). However, picornaviruses have been traditionally classified into several genera. The 5′ NTR sequences thought to be important for internal initiation of different genera are quite different and can be grouped together into three classes: (i) hepatitis A virus; (ii) cardioviruses and aphthoviruses, e.g., EMCV and foot-and-mouth disease virus; and (iii) enteroviruses and rhinoviruses, e.g., poliovirus and rhinovirus. As a result, the mechanism for initiation on cardiovirus and aphthovirus may be different from that for initiation on enterovirus and rhinovirus (19).

Swine vesicular disease virus (SVDV) is an enterovirus belonging to the family *Picornaviridae*. It causes an infectious disease in pigs that is characterized by the appearance of vesicles on the tongue, in the mouth, and on the feet and hocks (6, 25, 26). Like many picornaviruses, SVDV contains a very long 5′ NTR, 742 nucleotides in length, which contains nine potential translation start codons positioned before the codon that initiates the large open reading frame at nucleotide position 743 (18). Within the NTR, the predicted stem-loop structure, as described for the 5′ NTR of poliovirus, is conserved in SVDV (18, 19, 33). Therefore, it is very likely that the 5′ NTR of SVDV works as an IRES for protein translation.

In this study, we demonstrated the capability of internal initiation on the SVDV 5′ NTR through analysis of a bicistronic retroviral vector containing the sequence. In addition, we found that the internal initiation capability of the SVDV 5′ NTR was greatly enhanced in HeLa cells

* Corresponding author.

compared with that in NIH 3T3 cells. The expression efficiency of the IRES-based bicistronic retroviral vectors was also much higher than that of the conventional double-gene vector or the U3-based vector.

MATERIALS AND METHODS

Cell culture. All cell lines were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at 37°C in a 5% CO_2 incubator. This experiment used the GP+E-86 ecotropic retrovirus packaging cell line (22) and the GP+*env*Am12 amphotropic packaging cell line (23) to produce recombinant retroviruses. NIH 3T3 cells were the major target cells used for analysis in this study. Primary human skin fibroblast cells were isolated from foreskin biopsies of a normal human. Rat embryo fibroblast cells (REF) were isolated from 17- to 18-day-old Norway brown rat embryos by the cold-trypsin method (11).

Retroviral vector construction. All vectors were derived from the N2 vector (2). The Tn*5* neomycin resistance (Neo^r) gene and a cDNA that codes for porcine growth hormone (pGH), a gift from W. C. Chang (National Taiwan University), were used as the tested genes in constructions. This pGH cDNA includes the 5′ nontranslated region, the signal peptide region plus the full coding region of pGH, and the 3′ nontranslated region but excludes the polyadenylation signal. Plasmid P (see Fig. 1) was constructed by replacing the Neo^r gene with the pGH cDNA fragment through blunt-end ligation to the N2 vector at the *Eco*RI site (2). The NP and PN plasmids were constructed by placing the pGH cDNA downstream or upstream of the Neo^r gene by blunt-end ligation to the N2 vector at the *Xho*I or *Eco*RI site, respectively. Such constructions thus put two genes within the same transcriptional unit. A 742-bp-long 5′ NTR cDNA fragment of SVDV was obtained by reverse transcription-polymerase chain reaction amplification of the viral RNA, which was prepared from the virus particles isolated from the stool of SVDV-infected pigs (30) and was kindly provided by H. S. Lai (National Taiwan University). The primer sequences used (5′-TTAAAACAGCCTGTGGGTTG-3′ and 5′-TTTAACGTATTGAGCGTTAT-3′) were adopted from the published data (18). This segment contains the putative IRES but excludes the authentic AUG initiation codon of SVDV. Plasmids NSP and PSN were constructed by insertion of the 5′ NTR cDNA fragment of SVDV between the Neo^r gene and the pGH cDNA on plasmids NP and PN, respectively. Plasmid PMX contained an internal metallothionein (MT) promoter cloned at the *Eco*RI site of the N2 vector. The pGH cDNA was placed under control of the LTR promoter, while the Neo^r gene was under control of the MT promoter. The last plasmid, CB, is a U3-based retroviral vector. The pGH cDNA was first cloned downstream of the immediate-early gene promoter of cytomegalovirus (CMV); the CMV-pGH cassette was subsequently inserted into the 3′ U3 region at the *Bam*HI site before the LTR promoter (4). A unique *Hin*dIII site was created at the 3′ end of CMV-pGH and could be used to check the duplication event of the 3′ U3 region after reverse transcription (see Fig. 2).

Virus production and infection. Virus was generated by transfecting 2×10^5 ecotropic GP+E-86 or amphotropic GP+*env*Am12 packaging cells with 20 μg of vector DNA by using the calcium phosphate precipitation procedure (14). For the P plasmid, which has no Neo^r gene, 18 μg of P plasmid DNA was cotransfected with 2 μg of pSV2neo DNA (used as a selectable marker) into the packaging cell lines. Transfected cells were selected with 0.8 mg of G418 per ml, and resistant colonies were pooled to form the virus-producing cells. Viruses were harvested 16 to 18 h after fresh medium was placed onto the nearly confluent virus-producing cells and were used to infect NIH 3T3, primary human fibroblast, and REF target cells. Infection was performed in the presence of 8 μg of Polybrene per ml. Since NIH 3T3 cells infected with P virus were not selectable, they were coinfected with 2 ml of pGH retrovirus and 2 μl of Neo^r virus (i.e., N2 virus) to obtain G418-resistant clones which were simultaneously infected with the pGH retrovirus (17). Selection was continued until colonies appeared. All of the G418-resistant clones (>1,000) were pooled for further analysis.

Southern and Northern (RNA) blot analyses. High-molecular-weight DNA was digested with restriction enzymes, electrophoresed on a 0.8% agarose gel, and electrotransferred to a Nytran membrane (Schleicher & Schuell, Inc., Keene, N.H.). The pGH-specific probe was used to hybridize the membrane. RNA was extracted from cultured cells by the acid guanidinium thiocyanate method (5). Ten micrograms of total RNA was subjected to glyoxal-dimethyl sulfoxide-agarose gel electrophoresis (31), transferred to the Nytran membrane, and hybridized with the pGH probe.

ELISA for pGH. Rabbit anti-pGH serum was obtained by immunizing rabbits intrasplenically (16) and was kindly provided by W. C. Chang. In the enzyme-linked immunosorbent assay (ELISA), standard pGH (a gift from International Mineral Corp.) at several known concentrations and culture medium samples were first coated onto 96-well plates. Rabbit antiserum (1:10,000 dilution) was added and allowed to react with pGH at 37°C for 1 h. After removal of the first antibody by four washes with phosphate-buffered saline (PBS), a second antibody, goat anti-rabbit immunoglobulin antibody conjugated with horseradish peroxidase, was added (1:3,000 dilution) for a further 1 h of incubation. The plates were washed again thoroughly with PBS, and 100 μl of substrate solution (2 mg of *o*-phenylenediamine dissolved in 1 ml of 1 M phosphate citrate [pH 5.0] plus 0.02% H_2O_2) was added for development. After 30 min, 100 μl of 1 M H_2SO_4 was added immediately to stop the reaction and the colors were measured by optical density at 492 nm. A standard pGH curve based on the optical density at 492 nm of a known pGH concentration was plotted, and the concentrations of pGH in our samples were obtained by interpolation from the standard curve.

Western blot (immunoblot) analysis. To analyze the pGH protein translated from the infected clones, cells were grown in a 10-cm dish until confluent and washed twice with PBS. A serum-free medium, Iscove's modified Dulbecco's medium, was used to refresh the cells overnight. The overnight-culture medium was collected, concentrated by Amicon Centriprep-10, and subjected to sodium dodecyl sulfate–12.5% polyacrylamide gel electrophoresis (21). Proteins were transferred from the gel to a nitrocellulose filter (3). The pGH protein was reacted with rabbit anti-pGH serum and subsequently with an alkaline phosphatase-conjugated second antibody, goat anti-rabbit immunoglobulin G. The blot was then developed with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate for a few minutes until color developed.

Quantitative assay of NPTII activity. Quantitative assay of neomycin phosphotransferase (NPTII) was done as described by Reiss et al. (29). Infected cells were harvested and adjusted to equal intensities (5×10^5 cells per 20 μl for NIH 3T3 cells and 7.5×10^5 cells per 20 μl for HeLa cells) with TE buffer (10 mM Tris [pH 7.4], 1 mM EDTA). Crude cell

extract was obtained by sonication and separation from cellular debris by centrifugation. Twenty microliters of the extract was subjected to nondenaturing 10% polyacrylamide gel electrophoresis. The polyacrylamide gel was equilibrated in 100 ml of reaction buffer (67 mM Tris maleate [pH 7.1], 4.2 mM $MgCl_2$, 400 mM NH_4Cl) for 30 min, transferred onto a glass plate, and covered with a second gel made up of 1% agarose in the same reaction buffer containing the substrates (0.5 to 2 nM [γ-^{32}P]ATP and 20 μg of kanamycin sulfate per ml). After incubation for 30 min at room temperature, a sheet of Whatman P81 paper, two sheets of Whatman 3MM paper, a stack of blotting paper, and a 1-kg weight were put on top of the agarose gel. After 3 h, the P81 papers were removed, washed twice with 500 ml of hot water (80°C) and several times with cold water, dried, and exposed to X-ray film with an intensifying screen.

RESULTS

Efficient internal initiation and high viral titer from a 5′ NTR-containing bicistronic retroviral vector. The retroviral vectors used here were N2 vector derivatives (2). The LTR promoter was derived from the Moloney murine leukemia virus genome. The virus sequences retained in the vector consisted of the packaging signal and a pair of cryptic splicing signals. A full-length RNA that codes for an out-of-frame *gag* insert fusion protein, as well as a cryptically spliced mRNA that codes for the authentic protein of the inserted sequences, was detected (2). Four types of retroviral vectors were constructed (Fig. 1). The first type was a single-gene vector, P, which contained only the pGH cDNA in the retroviral vector and was not selectable. The second type was a bicistronic vector. NP and PN contained two exogenous genes, those for pGH and Neo^r, without the 5′ NTR segment of SVDV in the intercistronic region, whereas NSP and PSN contained a 742-bp fragment of the 5′ NTR of SVDV inserted between two cistrons. The third type was a conventional retroviral vector, PMX, which included a pGH cDNA driven by the LTR promoter and a Neo^r gene driven by an internal MT promoter within the LTR transcriptional unit. The last type was a U3-based retroviral vector, CB, in which the CMV-pGH expression cassette was inserted at the 3′ U3 region, thus allowing the cassette to be duplicated to the U3 region of the 5′ LTR after one cycle of reverse transcription (4, 15).

All of the plasmid DNAs except P were directly transfected into the GP+E-86 ecotropic packaging cell line, viruses collected from GP+E-86 cells were used to infect NIH 3T3 cells, and virus titers were determined. Since no selectable marker was included in vector P, the GP+E-86 cell line was cotransfected with 18 μg of vector DNA and 2 μg of pSV2neo DNA to generate a stable, virus-producing cell line. Similarly, NIH 3T3 cells were coinfected with P virus and Neo^r virus to produce a stable, G418-resistant cell line which simultaneously expressed pGH (17). Viruses with titers ranging around 10^5 CFU/ml, comparable to that of the virus from cells containing the N2 vector, were obtained from cells harboring vectors NP, NSP, and PSN, but only 10^3 CFU/ml was obtained from cells harboring vectors PMX and CB. No colonies of cells transfected or infected with the PN construct were observed, a result reproducibly obtained. The results therefore have two implications. (i) Expression of Neo^r gene from the downstream cistron of vector PN was not sufficient to confer G418 resistance on the GP+E-86 cell line. However, the 5′ NTR of SVDV contained in the intercistronic region of vector PSN immediately conferred sufficient G418 resistance to GP+E-86 and NIH 3T3 cells, suggesting that the 5′ NTR of SVDV could function as an IRES. (ii) Inclusion of the 742-bp SVDV 5′ NTR fragment within the retroviral vector did not adversely affect the virus titer, since the bicistronic vectors had titers similar to that of the single-gene vector. On the contrary, both the conventional retroviral vector and the U3-based vector had much lower titers, apparently owing to interference between two promoters (4, 10).

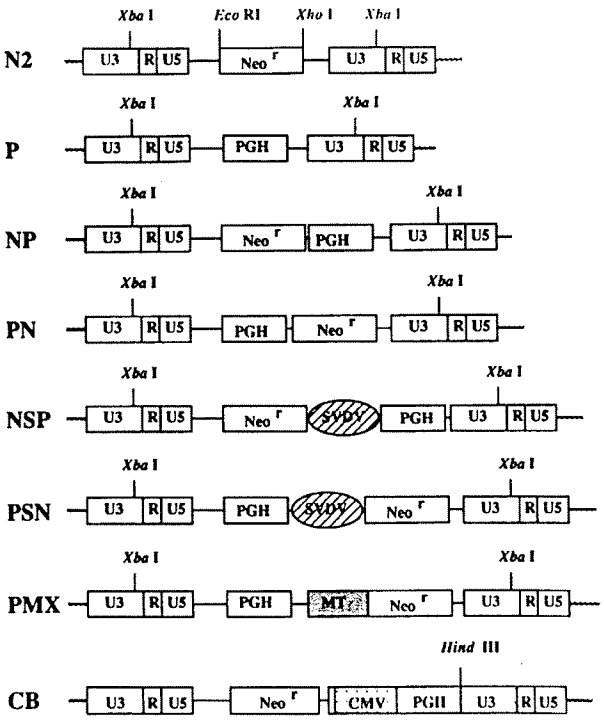


FIG. 1. Retroviral vectors. The N2 vector was described previously (2). All of the vectors shown were constructed on the basis of the backbone of N2. Abbreviations: PGH, the cDNA that codes for pGH; Neo^r, DNA for neomycin phosphotransferase; SVDV, the cDNA fragment of the 5′ NTR of SVDV; MT, the MT promoter; CMV, the immediate-early gene promoter of CMV. Names for the bicistronic vectors are based on the orders of DNA fragments. P stands for pGH, N stands for Neo^r, and S stands for the 5′ NTR of SVDV. PMX is a conventional retroviral vector with an MT-Neo^r cassette inserted at the *Xho*I site of N2; CB is a U3-based vector with a CMV-pGH cassette inserted at the *Bam*HI site of the 3′ U3 region of the N3 vector (4).

Expression of exogenous genes from the IRES-based bicistronic retroviral vector. The internal initiation capability of the 5′ NTR of SVDV was primarily demonstrated by the above-described experiment; we further performed detailed analysis to understand how it worked and whether it would affect the translation of the upstream cap-dependent cistron within the retroviral vector. To avoid bias due to the different expression efficiencies that resulted from different integration sites in individual clones, we pooled all of the G418-resistant colonies (>1,000) from each infection experiment for analysis. Figure 2A shows the results of Southern blot analysis of the proviral DNA present in the construct-infected clones. Restriction enzyme *Xba*I or *Hin*dIII was

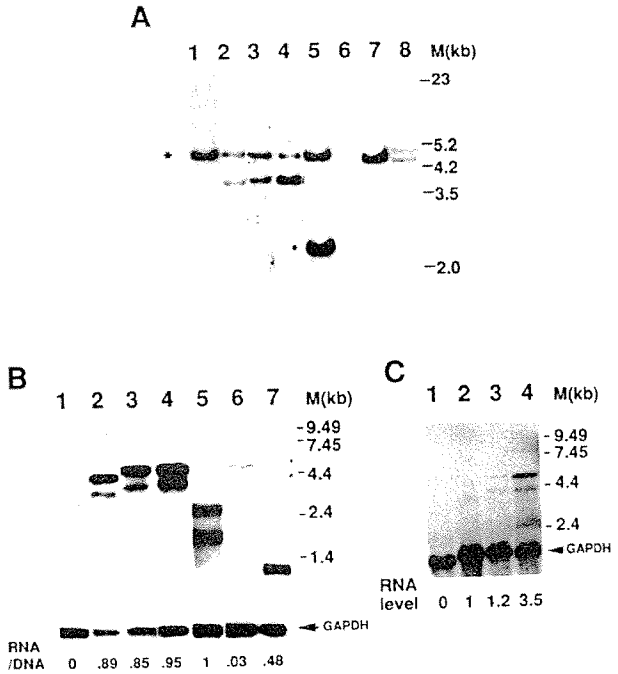


FIG. 2. Transcriptional rates of various retroviral vectors in infected cells. (A) Southern blot analysis of the proviral DNA in infected NIH 3T3 cells. Genomic DNAs extracted from infected cells were digested with either *Xba*I (lanes 1 to 6) or *Hin*dIII (lanes 7 and 8). Lanes 1 and 7 contained the NIH 3T3 cell control; the others contained cells infected with NP (lane 2), NSP (lane 3), PSN (lane 4), P (lane 5), PMX (lane 6), or CB (lane 8) virus, respectively. The asterisk indicates the position of the endogenous gene for mouse growth hormone, which was cross-hybridized with the pGH probe. M, markers. (B) Northern blot analysis of total RNA from virus-infected NIH 3T3 cells. RNA extracted from NIH 3T3 control cells (lane 1) or cells infected with NP (lane 2), NSP (lane 3), PSN (lane 4), P (lane 5), PMX (lane 6), or CB (lane 7) virus were analyzed. The filter was hybridized with a pGH probe and the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) probe, which was used as an internal control and whose position is indicated. Transcriptional rates were estimated by first adjusting the RNA level with the internal GAPDH control and then dividing the resulting RNA amount with the average DNA copy number shown in panel A, based on densitometric quantitation, and is expressed as the RNA/DNA ratio shown at the bottom of each lane. (C) Northern blot analysis of total RNA from infected HeLa cells. Lane 1 contained the HeLa cell control; the others contained cells infected with NP (lane 2), NSP (lane 3), or PSN (lane 4) virus. The relative RNA levels normalized to the GAPDH level are shown at the bottom.

used to detect the integrity of the proviral DNA. The results showed that there was no gross rearrangement of the proviral genome in any of the infected clones. Duplication of the chimeric U3 region was observed (Fig. 2A, lane 8) in CB virus-infected NIH 3T3 cells, as described previously (4). When normalized to the endogenous gene of the mouse growth hormone (indicated by an asterisk) which was cross-hybridized with the pGH probe, most of the proviral DNAs were present as one copy, on average, in the genome of the infected cells, except for PSN and P, which were present as two copies (lanes 4 and 5). Northern blot analysis of the RNA expressed from each vector is shown in Fig. 2B. All of the clones harboring constructs NP (lane 2), NSP (lane 3), PSN (lane 4), and P (lane 5) contained the full-length RNA,

TABLE 1. pGH production in infected cells

Vector	pGH production (ng/10^6 cells/day)[a]			
	NIH 3T3	HeLa	REF	Human skin fibroblast
NP	—[b]	—	ND[c]	ND
NSP	24	40	ND	ND
PSN	744[d]	40[d]	1,180	760
P	680[d]	ND	ND	ND
PMX	13	ND	ND	ND
CB	362	ND	ND	ND

[a] Overnight-culture medium was collected, and pGH protein was quantitated by ELISA (see Materials and Methods).
[b] —, the level of pGH protein was too low to be detected by ELISA.
[c] ND, not done.
[d] The level of pGH protein was divided by two since there are two copies of proviral DNA in the genome of this infected clone.

as well as the cryptically spliced RNA, characteristic of the N2 vector (2). The expression efficiency of vector NP, NSP, or PSN, after being adjusted with respect to the proviral DNA copy number, was equivalent to that of single-gene vector P, a result consistent with the viral titers obtained. However, the pGH RNA expressed from the PMX vector was dramatically reduced (lane 6), to only about 3% of that of the P vector. The pGH RNA expressed from the CMV promoter of the CB vector was about half of that of the P vector (lane 7). In cells containing the CB vector, the full-length RNA derived from the LTR promoter was strongly suppressed by the CMV promoter; thus, it was detectable only in a longer exposure of the Northern blot (data not shown). These data confirmed the problem of promoter interference inherent in the double-gene vectors, e.g., PMX and CB, and demonstrated the advantage of the IRES-based bicistronic vector.

The pGH protein translated from the mRNA derived from each vector was analyzed by ELISA and Western blotting, and the results are shown in Table 1 and Fig. 3A, respectively. As expected, the amount of pGH protein produced from the NP vector was too small to be detected (lane 2). On the contrary, pGH protein was produced from both NSP and PSN vectors but was recovered in 60-fold excess from the latter (lanes 3 and 4, respectively). If the RNA amount of individual clones was taken into consideration, the translational efficiency of the unselectable pGH gene from the upstream cistron was 30-fold higher than that from the downstream cistron (Fig. 3A and Table 1). Nevertheless, when the enzymatic activity of NPTII was measured, protein translation from the downstream cistron was almost as efficient as that from the upstream cistron after adjustment of the protein activity for the RNA amount (Fig. 3B, lanes 3 and 4). On the other hand, the translational efficiency of pGH protein from double-gene vector PMX or CB was more or less comparable to that from vector P after adjustment for the RNA level of each clone (Fig. 3A and Table 1).

These data therefore suggested that insertion of the SVDV IRES in the bicistronic retroviral vector did not affect the cap-dependent translation of the first cistron; the internal initiation efficiency from the SVDV IRES was only about 1/30 of that of the cap-dependent translation when the gene was not selectable but was increased to an equivalent level after the gene was selected.

Stimulation of the internal initiation activity of the SVDV IRES in HeLa cells. By using the IRES derived from EMCV in a bicistronic retroviral vector, Adam et al. (1) have

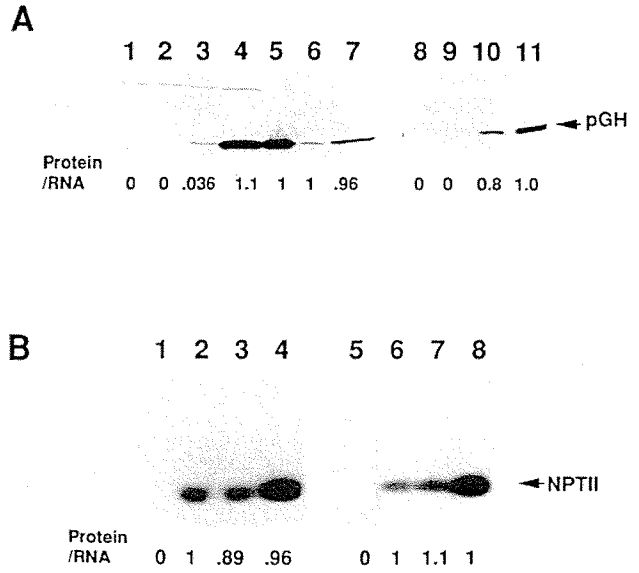


FIG. 3. Quantitation of protein expression from the retroviral vectors. (A) Western blot analysis of the pGH protein expressed from each vector. The pGH protein secreted into the supernatant of cultured cells was collected and analyzed by Western blotting (see Materials and Methods). The orders for lanes 1 to 7 and 8 to 11 are the same as those described in the legend to Fig. 2B and C, respectively. The position of the 22-kDa pGH protein is indicated. The translational efficiency of pGH from each vector (or cistron) is represented by the protein/RNA ratio shown at the bottom. (B) Assay of NPTII activity. The crude extracts of equal cell numbers (5×10^5 for NIH 3T3 cells and 7.5×10^5 for HeLa cells) were fractionated on a nondenaturing polyacrylamide gel. NPTII activity was assayed by in situ reaction of the polyacrylamide gel with an agarose gel containing the substrates kanamycin and [γ-^{32}P]ATP (see Materials and Methods). The position of the NPTII enzyme is indicated. Quantitation of enzyme activity was done by cutting out the spot from P81 paper and counting the radioactivity with a scintillation counter. The translational efficiency of NPTII is represented by the protein/RNA ratio shown at the bottom. Lanes 1 to 8 depict control cells (lanes 1 and 5) and NP (lanes 2 and 6), NSP (lanes 3 and 7), and PSN (lanes 4 and 8) virus-infected NIH 3T3 and HeLa cells, respectively.

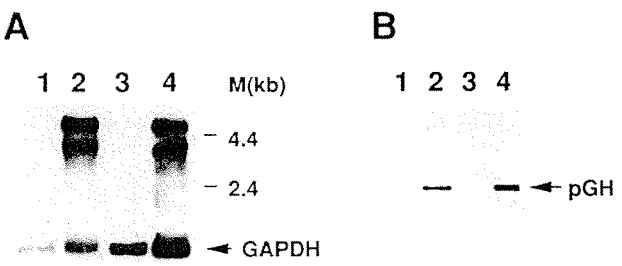


FIG. 4. Expression of the bicistronic retroviral vector in transduced primary cells. Expression of pGH in PSN virus-infected primary REF and human skin fibroblast cells was analyzed either by Northern blotting with the pGH probe (A) or by Western blotting with rabbit anti-pGH serum (B). Lanes: 1, human fibroblast cell control; 2, PSN virus-infected human fibroblast cells; 3, REF cell control; 4, PSN virus-infected REF. M, markers; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

previously shown that the downstream coding region was expressed at a fourfold lower level than that directed by the first cistron. This is in contrast to our finding of a 30-fold difference (Fig. 3A). Because the IRES sequence of EMCV is quite different from that of SVDV, the mechanisms of internal initiation of the two viruses could be quite different (19). It has been demonstrated that some specific cellular factors are present in HeLa cell extract that can stimulate the IRES function of poliovirus (7). It was therefore interesting to determine whether this would apply to the SVDV IRES and whether this may account for the dramatic reduction of translational efficiency of the internal initiation in our bicistronic vectors. NP, NSP, and PSN viruses were therefore used to infect HeLa cells. The RNA and protein were analyzed as described above, and the results are shown in Fig. 2C and 3, respectively. Interestingly, the pGH protein initiated from the IRES in the NSP vector was greatly increased relative to that from the first cistron of the PSN vector (Fig. 3A, lanes 10 and 11 versus lanes 3 and 4). Considering the different RNA amounts in the clones (Fig. 2C, lanes 3 and 4), the translational efficiency derived from the SVDV IRES was increased to a level close to that of the cap-dependent translation of the first cistron in HeLa cells. These results indicated that there was some specific cellular factor(s) in HeLa cells that could stimulate the internal initiation activity of the SVDV IRES. However, the stimulation effect was not observed when NPTII activities were compared between HeLa cells harboring the PSN vector and cells harboring the NSP vector (Fig. 3B, lanes 7 and 8 versus lanes 3 and 4).

Analysis of expression of the bicistronic vector in primary cells. To determine how the bicistronic vector is expressed in primary cells, REF and human skin fibroblast cells were infected with PSN virus. In both cases, the recombinant retrovirus infected the primary cells efficiently. Expression of pGH was analyzed at RNA and protein levels. Results showed that RNA was highly transcribed from the retroviral LTR promoter with the expected pattern in both fibroblast cells (Fig. 4A). The cap-dependent translation of pGH was also not affected by the presence of the SVDV IRES (Fig. 4B and Table 1). The secretion rates of pGH from REF and human skin fibroblast cells were estimated to be 1.18 and 0.76 $\mu g/10^6$ cells per day, respectively (Table 1).

DISCUSSION

The major concern of coexpressing two genes in a retroviral vector is often poor expression of one of them. This is due mainly to promoter interference (10), since most retroviral vectors contain two promoters for this purpose. Although the U3-based vector has been designed to overcome this problem (4, 15), promoter interference on the vector DNA still remains in the DNA-transfected packaging cells, resulting in lower titers of the recombinant viruses. The results we obtained strongly confirmed this notion. The discovery of the internal-initiation property of the 5′ NTR of picornavirus has allowed the translation of two genes from a single transcript (1, 12). Such a construct, as only one promoter is included, is devoid of promoter interference. In this study, we adopted the IRES derived from the SVDV genome to construct a bicistronic retroviral vector and successfully demonstrated its internal initiation capability. These results corroborated those of previous studies, in which an EMCV IRES was used (1, 12). Most importantly, the overall coexpression efficiency from the IRES-based vector was much higher than that from the conventional retroviral vector or from the U3-based vector (Fig. 3A). As

shown in Fig. 2B, the IRES-based vector was much more advantageous than the other two types of vectors, basically at the transcriptional level. Otherwise, the translational efficiency of pGH from the upstream position of the IRES-based vector was comparable to that from the LTR-driven cistron (vector PMX) or from the CMV-driven cistron (vector CB), representative of conventional vectors or U3-based vectors, respectively (4).

That our bicistronic vector differed from previous ones (1) in the difference of translational efficiency between two cistrons was probably a result of the different IRES sources used. Although both EMCV and SVDV are picornaviruses, they belong to different subgroups. It has been reported that cardiovirus-aphthovirus and enterovirus-rhinovirus are distinctly different not only in IRES sequences but also in translation characteristics (19). For example, foot-and-mouth disease virus and EMCV RNAs are both translated with very high efficiency and high accuracy in the rabbit reticulocyte lysate. On the other hand, poliovirus is translated very inefficiently and inaccurately in the rabbit reticulocyte lysate but efficiently and accurately in HeLa or L-cell extracts (27). Addition of HeLa extracts to reticulocyte lysate increases the fidelity and efficiency of poliovirus RNA translation (7), implying that specific cellular factors present in HeLa cells are required for efficient initiation on poliovirus RNA. Such a cellular factor, a 57-kDa protein, was recently identified from HeLa cell extracts (28). Because SVDV is an enterovirus, translation initiation is more similar to that of poliovirus. Indeed, we found that the translational efficiency of pGH from the IRES-initiated cistron was greatly increased when vectors were transduced into HeLa cells (Fig. 3A), implying that the cellular factors were cell type specific. However, when a Neo^r gene was assayed, it was found to be translated with equivalent efficiency from either the LTR-directed cistron or the IRES-directed cistron in both NIH 3T3 and HeLa cells. No significant stimulation of internally initiated NPTII activity was observed in HeLa cells either. The different results obtained with regard to pGH protein and NPTII enzymatic activity may be due to the selection pressure exerted on the infected clones. As a result, NPTII activity was forced to increase to a level sufficient to render cells viable; the activity measured may not reflect a real translational effect of different cistrons. Alternatively, the translational effect may vary from target gene to target gene, although we think this is less likely. To exclude the latter possibility, more nonselectable genes could be tested with this bicistronic vector in both NIH 3T3 and HeLa cells. We are currently performing this experiment.

In conclusion, our study suggests that the IRES-based vector is a better retroviral vector for coexpression of two genes with higher efficiency. Because the retroviral vector transduces genes into the chromosome through its unique integration mechanism, the RNA transcribed from an IRES-based vector is derived solely from the LTR promoter and no internally initiated mRNA could account for translation of the second cistron (1, 12; Fig. 2B). This system is therefore a powerful tool with which to study the internal initiation mechanism in a stable cell line and to identify any other potential IRES. It is also useful for studying the different mechanisms of internal initiation employed by different picornaviruses.

ACKNOWLEDGMENTS

We thank H. S. Lai for SVDV viral RNA and W. C. Chang for pGH cDNA and rabbit anti-pGH serum.

This work was supported by grant NSC 82-0418-B-002-098-BC from the National Science Council.

REFERENCES

1. **Adam, M. A., N. Ramesh, A. D. Miller, and W. R. A. Osborne.** 1991. Internal initiation of translation in retroviral vectors carrying picornavirus 5′ nontranslated regions. J. Virol. **65:** 4985–4990.
2. **Armentano, D., S. F. Yu, P. W. Kantoff, T. von Ruden, W. F. Anderson, and E. Gilboa.** 1987. Effect of internal viral sequences on the utility of retroviral vectors. J. Virol. **61:**1647–1650.
3. **Burnette, W. H.** 1981. Electrophoretic transfer of proteins from SDS-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. **112:**195–203.
4. **Chen, B. F., C. L. Hsieh, D. S. Chen, and L. H. Hwang.** 1992. Improved gene expression by a U3-based retroviral vector. Biochem. Biophys. Res. Commun. **184:**330–337.
5. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156–159.
6. **Dawe, P. S., A. J. Forman, and C. J. Smale.** 1973. A preliminary investigation of the swine vesicular disease epidemic in Britain. Nature (London) **241:**540–542.
7. **Dorner, A. J., B. L. Semler, R. J. Jackson, R. Hanecak, E. Duprey, and E. Wimmer.** 1984. In vitro translation of poliovirus RNA: utilization of internal initiation sites in reticulocyte lysate. J. Virol. **50:**507–514.
8. **Eglitis, M. A., and W. F. Anderson.** 1989. Retroviral vectors for introduction of genes into mammalian cells. BioTechniques **6:**608–614.
9. **Emerman, M., and H. M. Temin.** 1984. Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. Cell **39:**459–467.
10. **Emerman, M., and H. M. Temin.** 1986. Quantitative analysis of gene expression in integrated retroviral vectors. Mol. Cell. Biol. **6:**792–800.
11. **Freshney, R. I.** 1983. Culture of animal cells: a manual of basic technique, p. 107–110. Alan R. Liss, Inc., New York.
12. **Ghattas, I. R., J. R. Sanes, and J. E. Majors.** 1991. The encephalomyocarditis virus internal ribosome entry site allows efficient coexpression of two genes from a recombinant provirus in cultured cells and in embryos. Mol. Cell. Biol. **11:**5848–5859.
13. **Gilboa, E., M. A. Eglitis, P. W. Kantoff, and W. F. Anderson.** 1986. Transfer and expression of cloned genes using retroviral vectors. BioTechniques **4:**504–512.
14. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52:**456–467.
15. **Hantzopoulos, P. A., B. A. Sullenger, G. Ungers, and E. Gilboa.** 1989. Improved gene expression upon transfer of the adenosine deaminase minigene outside the transcriptional unit of a retroviral vector. Proc. Natl. Acad. Sci. USA **86:**3519–3523.
16. **Hong, T. H., S. T. Chen, T. K. Tang, S. C. Wang, and T. H. Chang.** 1989. The production of polyclonal and monoclonal antibodies in mice using novel immunization methods. J. Immunol. Methods **120:**151–157.
17. **Hwang, L. H., B. F. Chen, P. J. Lee, S. Y. Ho, and J. J. Liu.** 1992. Use of helper-free retroviral vector to direct a high expression of porcine growth hormone in mouse fibroblast cells. Biotech. Appl. Biochem. **16:**171–181.
18. **Inoue, T., T. Suzuki, and K. Sekiguchi.** 1989. The complete nucleotide sequence of swine vesicular disease virus. J. Gen. Virol. **70:**919–934.
19. **Jackson, R. J., M. T. Howell, and A. Kaminiski.** 1990. The novel mechanism of initiation of picornavirus RNA translation. Trends Biochem. Sci. **15:**477–483.
20. **Kozak, M.** 1989. The scanning model of translation: an update. J. Cell Biol. **108:**229–241.
21. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
22. **Markowitz, D., S. Goff, and A. Bank.** 1988. A safe packaging

line for gene transfer: separating viral genes on two different plasmids. J. Virol. **62**:1120–1124.
23. **Markowitz, D., S. Goff, and A. Bank.** 1988. Construction and use of a safe and efficient amphotropic packaging cell line. Virology **167**:400–406.
24. **Miller, M. D., and G. J. Rosman.** 1989. Improved retroviral vectors for gene transfer and expression. BioTechniques **7**:980–990.
25. **Mowat, G. N., J. H. Darbyshire, and J. F. Huntley.** 1972. Differentiation of a vesicular disease of pigs in Hong Kong from foot-and-mouth disease. Vet. Rec. **90**:618–621.
26. **Nardelli, L., E. Lodetti, G. Gualandi, R. Burrows, D. Gordridge, F. Brown, and B. Cartwright.** 1968. A foot and mouth disease syndrome in pigs caused by an enterovirus. Nature (London) **219**:1275–1276.
27. **Pelletier, J., G. Kaplan, V. R. Racaniello, and N. Sonenberg.** 1988. Translational efficiency of poliovirus mRNA: mapping inhibitory *cis*-acting elements within the 5′ noncoding region. J. Virol. **62**:2219–2227.
28. **Pestova, T. V., C. U. T. Hellen, and E. Wimmer.** 1991. Translation of poliovirus RNA: role of an essential *cis*-acting oligopyrimidine element within the 5′ nontranslated region and involvement of a cellular 57-kilodalton protein. J. Virol. **65**:6194–6204.
29. **Reiss, B., R. Sprengel, H. Will, and H. Schaller.** 1984. A new sensitive method for qualitative and quantitative assay of neomycin phosphotransferase in crude cell extracts. Gene **30**:211–218.
30. **Rotbart, H. A.** 1990. PCR amplification of enterovirus, p. 372–377. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., San Diego, Calif.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., vol 1, p. 7.40–7.42. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Temin, H. M.** 1989. Retroviral vectors: promise and reality. Science **246**:983.
33. **Tracy, S., H.-L. Liu, and N. M. Chapman.** 1985. Coxsackievirus B3: primary structure of the 5′ noncoding and capsid protein-coding regions of the genome. Virus Res. **3**:263–270.